DOCUMENT RESUME

ED 299 304 TM 012 098

AUTHOR Kane, Michael
TITLE Testing New Teachers: A Report of a Review of Teacher Testing Practices and Materials in the District of Columbia Public Schools.
INSTITUTION District of Columbia Public Schools, Washington, DC. Div. of Quality Assurance and Management Planning.
PUB DATE Jan 88
NOTE 146p.; For a related document, see TM 012 099.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Beginning Teachers; Elementary Secondary Education; *Licensing Examinations (Professions); Material Development; *Public School Teachers; School Policy; Standards; *Teacher Certification; Teacher Evaluation; Teacher Qualifications; *Test Construction; Testing Programs; Test Validity
IDENTIFIERS *District of Columbia Public Schools; *Teacher Competency Testing

ABSTRACT

The current status of teacher testing practices and materials in the public schools of the District of Columbia was evaluated. This document reports the findings and recommendations of eight professionals who convened in Washington, D.C., on November 17-19, 1987. These consultants, via interviews with practitioners, found a high degree of readiness within the district for the use of subject matter examinations as a criterion for teacher certification. Issues were examined by a policy analysis tool known as the convening process. Recommendations for a teacher testing policy are provided: (1) subject matter knowledge testing for teacher certification or licensure; (2) requirement of a specified certification score; (3) certification testing for all teachers regardless of other certification; (4) limitations on temporary certification; (5) analysis of processes used in hiring, promotion, and tenure decisions; (6) review of tests and development of new tests; and (7) a writing test for diagnostic and prescriptive purposes for all new employees. Specific recommendations were made to improve the current validity procedures and test development process, including the establishment of a research and measurement unit to collaborate with the current Intern-Mentor program. An overview of legal issues in teacher testing, a report on test development and intern performance, a consultant briefing paper, meeting agenda and a list of interviewees, and sample revalidation procedures and forms are appended. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299304

TESTING NEW TEACHERS

A REPORT OF A REVIEW OF TEACHER TESTING PRACTICES AND MATERIALS IN THE DISTRICT OF COLUMBIA PUBLIC SCHOOLS

DIVISION OF QUALITY ASSURANCE

AND

MANAGEMENT PLANNING

DAVID L. HUIE
DIRECTOR

RESEARCH AND EVALUATION BRANCH

NORMAN GOLD
DIRECTOR

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

JANUARY 1988

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MICHAEL KANE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 012 098

BEST COPY AVAILABLE

PREFACE

From November 17th to 19th, 1987 a small group of professionals was convened by the Division of Quality Assurance to examine the current teacher testing practices and materials of the District of Columbia Public Schools. These individuals, from state departments of education, universities and a law firm, were invited to participate because of their state-of-the-art experience either designing, implementing, studying or defending the use of teacher tests as a screening device for the certification and/or selection of school teachers. The questions put to the group were simple and straight forward: What would be an appropriate teacher testing policy for the District of Columbia Public Schools given its particular personnel needs and status? What is the current state of development of its existing teacher tests relative to the state-of-the-art? What is the best use for the existing tests in light of the responses to the first two questions? And, how should the district proceed to implement the recommended teacher testing policy?

The process used to examine the issues presented the District's consultants was based on a model developed by Norman Gold, Director of Research and Evaluation, while a Senior Research Associate at the National Institute of Education. The "Convening Process", as it is known, has been used as a policy analysis tool in several locations to examine significant, pressing issues of educational policy. It has been used twice previously in the District of Columbia Public Schools. Barbara Williams and Michael Kane organized and co-chaired the process, a full description of which can be found in Appendix C.

This is a report of the findings and recommendations of the convening process. All participants authored sections of the report. Michael Kane, in his role as chair, authored this final version of the report.

PARTICIPANTS

Dr. Michael Kane, Chair
President, MCK Associates
Annapolis, Maryland

Dr. Barbara Williams, Co Chair
Independent Consultant
Washington, D.C.

Mr. David Boyd, Esquire
Partner, Balch and Bingham
Montgomery, Alabama

Dr. Ronald Braithwaite
Executive Director, Norfolk Area
Health Education Center
Norfolk, Virginia

Dr. Norman Gold
Director, Research and Evaluation
District of Columbia
Public Schools
Washington, D.C.

Dr. Dennis Holmes
Consultant, District of
Columbia Public Schools
Washington, D.C.

Dr. Sylvia Johnson
Professor, Howard University
Washington, D.C.

Dr. William Mehrens
Professor, Michigan State
University
East Lansing, Michigan

Dr. Vana Meredith-Dabney
Supervisor of Assessment
South Carolina Department of
Education
Columbia, South Carolina

Dr. Lester Solomon
Director, Teacher Assessment
Georgia Department of
Education
Atlanta, Georgia

EXECUTIVE SUMMARY

Since the 1985-1986 school year all beginning teachers hired by the District of Columbia Public Schools have participated in a year long period of supervised induction to professional practice. This internship period was established by the Board of Education in response to a recommendation of the 1984 Study of Teacher Incentives conducted by the District as part of a Congressional mandate to explore promising alternatives to reward exemplary teachers.

The study found that a significant percentage of teachers would be retirement eligible by the early 1990s and, therefore, recommended the District place a special focus on its recruitment and retention practices. Specific recommendations included the development of the internship as well as the use of subject matter tests for certification and/or selection purposes.

Shortly after the initiation of the District's Intern-Mentor Program, in December 1985, the Program contracted for the development of a series of tests of subject matter knowledge to be used as a component of the Intern-Mentor Program. Twelve subject matter knowledge tests and a writing test have been developed to date.

The Board of Education has expressed its intent to expand the role of teacher testing in the District. It has asked that the Superintendent report on the current test development effort being conducted by the Intern-Mentor Program and prepare, by February, 1988, a set of implementation and policy recommendations to support the implementation of a broader teacher testing program in DCPS.

This document represents one component of the efforts supporting the preparation the Superintendent's response. It reports the findings and recommendations of eight professionals who convened in Washington on November 17 - 19 1987, interview staff and community members and review the current status of the DCPS teacher testing practices and materials.

In general, the consultants found a high degree of readiness within the District for the use of subject matter examinations as a criterion for teacher certification. Most individuals interviewed supported the use of teacher testing. However, they also felt testing should not be the sole criterion for certification and that it should be utilized in a manner which recognized that tests, in and of themselves, are not the sole determinant of a teacher's competency. Seven specific recommendations for a teacher testing policy in DCPS are offered:

1. subject matter knowledge testing of teachers, aligned, to the extent possible, with existing certification/endorsement areas, be used in DCPS for purposes of certification or licensure.

2. in order to be granted a standard teaching certificate in DCPS, a candidate must achieve a specified score on the certification subject-matter test.

3. the certification test requirement apply to all persons seeking initial certification in DCPS, including persons previously certified elsewhere and teachers seeking additional endorsements.

4. to accommodate test administration constraints:

a. certification candidates who have not achieved the specified standard on the certification test, either because of failure or lack of opportunity to be tested, be granted a time limited exemption from the certification test requirement during which time the candidate, if otherwise qualified, would be granted a temporary certificate and permitted, if hired, to teach in DCPS.

b. certification candidates actually teaching without first having taken the certification test be required to take the test at the first available administration after their employment.

c. DCPS develop appropriate modifications to teacher contract forms to be used for candidates teaching without having passed the test, in order to take into consideration remedial, administrative and legal ramifications of the arrangement.

5. DCPS undertake an analysis of its current appraisal processes used in hiring, promotion and tenure decisions and develop an overall evaluation system that better supports either continued professional development and/or non-renewal and discharge decisions.

6. the certification testing program include the tests currently in development. For areas in which no test is being developed, options which should be considered include: 1. no test be developed, particularly in low incidence fields; 2. validation of existing tests (e.g. NTE); and 3. development of additional tests through the process currently being utilized.

7. the writing test not be a certification requirement, but be required of all newly-hired employees and used for diagnostic-prescriptive purposes.

To effectively support the implementation of these recommended policies, certain test development practices and procedures must be followed. Testing in situations where either licensure or employment can be denied is fraught with legal ramifications. Very specific test development practices must be followed in order to demonstrate a test is relevant to the purpose for which it is utilized. It is being recommended DCPS

expand its current testing practice to one which makes testing a requirement for licensure.

Accordingly, a review of the test development procedures used to date was conducted to ascertain the developer's level of compliance with the standards of licensure testing. That review suggests that: 1. there were practices followed which are not sufficiently equivalent to the standards for licensure purposes and 2. there remain some test development functions, such as standard setting, to be completed. Further test development actions must, therefore, be undertaken. Nine specific recommendations are offered:

1. the Superintendent request the contractor to immediately inventory and submit all available documentation to DCPS for review.

2. DCPS, with outside consultation, conduct further technical and legal reviews of this documentation to guide future test development.

3. further development be undertaken and accompanied by written documentation to assure all aspects of test quality.

4. although DCPS and the contractor developed the tests using a statement of purpose which appears consistent with the use of the test for initial certification, a more precise statement of the purpose be used in the collection of additional validity evidence.

5. the validation of the existing tests continue and include additional input from teachers, university specialists and measurement specialists.

6. further field testing be conducted in areas with sufficient numbers of applicants to have meaningful data for further refinement of tests and possibly for standard setting.

7. standard setting be conducted utilizing a multi-level approach after the first real test administration using impact data from the test results and possibly from the field tests.

8. It is recommended that, given the technical nature of the test development process, the further development of the DCPS tests be assigned to a unit with appropriate research and measurement expertise, while maintaining continued collaboration of the Intern-Mentor Program. This includes supervision of the NES contract and coordination of any external monitoring.

9. It is recommended there be established an ongoing external technical advisory committee.

I. INTRODUCTION

The District of Columbia Public Schools (DCPS) is developing a series of initiatives designed to improve the quality of instruction in its schools. Facing significant rates of turnover among its instructional personnel over the next decade, the District has directed the focus of much of this improvement initiative at its teacher recruitment, selection and induction practices. Its lead program in this area is the Intern-Mentor Program.

The Intern-Mentor program is designed to provide a period of supervised induction to practice for beginning teachers. This program was implemented in the 1985-86 school year under a Board of Education mandate that all new teachers entering the school system serve a year-long internship under the guidance of an experienced mentor teacher. The goal of the Program is "to promote professional growth and development by utilizing exemplary teachers to provide intensive assistance and guidance to new teachers...for the purpose of improving teaching and learning in the classroom" (Toward Excellence in Teaching, Intern-Mentor Program, 1985.)

An "important tool in this process...is the Intern Content Knowledge Assessment Program. Under this program interns are required to take a content knowledge test in their teaching fields. These tests help the intern and the mentor teacher in evaluating subject-matter strengths and weaknesses and in determining areas for further study or concentrated guidance" (Intern Content Knowledge Assessment Program, Descriptive Brochure, 1987.)

The DCPS Board of Education has expressed its intent to expand the role of teacher testing in the school system. The Superintendent has been directed to report on the current test development effort being conducted by the Intern-Mentor Program and prepare, by February, 1988, a set of implementation and policy recommendations to support the implementation of a broader teacher testing program in DCPS.

Given the complexity and significance of an expanded teacher testing program a special task force of representatives of the Divisions of Human Resource Management, Quality Assurance and the Office of Incentives programs was convened to help prepare a response to the Board's request for guidance on this matter. After engaging in a set of internal review activities concerning the procedures and interim outcomes of the DCPS test development process, as well as an examination of the legal requirements associated with utilizing tests in employment related situations, the task force decided to augment its investigation with an external review.

On November 17th through 19th 1987 the participants identified above convened in Washington D.C. to review the DCPS test development process and interview staff and community members concerning teacher testing in the school system. In response to our charge, we are recommending a series of policies and procedures to follow in implementing a teacher testing program for DCPS which is consistent with the state-of-the-art and practice. This document reports on our deliberations by presenting our findings concerning the current test development status of the district and our recommendations for future policy and test development actions. Appended to the report is a complete description of the process we followed in arriving at these findings and recommendations.

Before presenting the findings and recommendations some background on current practice in testing and employment in DCPS and elsewhere should be reviewed. Therefore section II contains a brief description of these relevant practices. It is supplemented by appendix A which provides an overview of legal issues in teacher testing. We recommend this appendix also be read before turning to the findings and recommendations. Section III contains findings and recommendations concerning policy development. Section IV presents the findings and recommendations concerning test development.

I. INTRODUCTION

The District of Columbia Public Schools (DCPS) is developing a series of initiatives designed to improve the quality of instruction in its schools. Facing significant rates of turnover among its instructional personnel over the next decade, the District has directed the focus of much of this improvement initiative at its teacher recruitment, selection and induction practices. Its lead program in this area is the Intern-Mentor Program.

The Intern-Mentor program is designed to provide a period of supervised induction to practice for beginning teachers. This program was implemented in the 1985-86 school year under a Board of Education mandate that all new teachers entering the school system serve a year-long internship under the guidance of an experienced mentor teacher. The goal of the Program is "to promote professional growth and development by utilizing exemplary teachers to provide intensive assistance and guidance to new teachers...for the purpose of improving teaching and learning in the classroom" (Toward Excellence in Teaching, Intern-Mentor Program, 1985.)

An "important tool in this process...is the Intern Content Knowledge Assessment Program. Under this program interns are required to take a content knowledge test in their teaching fields. These tests help the intern and the mentor teacher in evaluating subject-matter strengths and weaknesses and in determining areas for further study or concentrated guidance" (Intern Content Knowledge Assessment Program, Descriptive Brochure, 1987.)

The DCPS Board of Education has expressed its intent to expand the role of teacher testing in the school system. The Superintendent has been directed to report on the current test development effort being conducted by the Intern-Mentor Program and prepare, by February, 1988, a set of implementation and policy recommendations to support the implementation of a broader teacher testing program in DCPS.

Given the complexity and significance of an expanded teacher testing program a special task force of representatives of the Divisions of Human Resource Management, Quality Assurance and the Office of Incentives programs was convened to help prepare a response to the Board's request for guidance on this matter. After engaging in a set of internal review activities concerning the procedures and interim outcomes of the DCPS test development process, as well as an examination of the legal requirements associated with utilizing tests in employment related situations, the task force decided to augment its investigation with an external review.

On November 17th through 19th 1987 the participants identified above convened in Washington D.C. to review the DCPS test development process and interview staff and community members concerning teacher testing in the school system. In response to our charge, we are recommending a series of policies and procedures to follow in implementing a teacher testing program for DCPS which is consistent with the state-of-the-art and practice. This document reports on our deliberations by presenting our findings concerning the current test development status of the district and our recommendations for future policy and test development actions. Appended to the report is a complete description of the process we followed in arriving at these findings and recommendations.

Before presenting the findings and recommendations some background on current practice in testing and employment in DCPS and elsewhere should be reviewed. Therefore section II contains a brief description of these relevant practices. It is supplemented by appendix A which provides an overview of legal issues in teacher testing. We recommend this appendix also be read before turning to the findings and recommendations. Section III contains findings and recommendations concerning policy development. Section IV presents the findings and recommendations concerning test development.

II BACKGROUND

As described in the introduction, the Intern-Mentor Program is the required induction process for all inexperienced or "beginning" teachers hired by DCPS. Over the three years of the program's existence approximately four hundred-fifty interns have participated in the program. The development of this program has also focussed attention on the recruitment process itself.

Prior to the mid 1980s, DCPS, like most non-sun-belt school districts, had been undergoing enrollment declines. As a result, recruitment processes required little attention. With the increased rates of turnover, decreasing supply of new teachers and program improvement oriented policy actions of the second half of this decade, recruitment functions took on renewed importance.

Working together the Division of Human Resources and the Intern-Mentor Program have been developing a more focused recruitment process. A comprehensive system for projecting vacancies earlier in the school year and for keeping track of applicant files has been developed. New methods for announcing vacancies and sending recruitment teams to major teacher education institutions have been established. Greater scrutiny of factors affecting recruiting effectiveness such as salary and residency considerations is resulting in more timely feedback to the system concerning needed policy actions.

Once recruited, prospective teachers must seek certification to teach in DCPS. The certification criteria are based primarily on the completion of required coursework in an accredited institution of higher education. Three classes of certification are given: standard; provisional, for those with minor deficiencies in their preparation program; and temporary, for those without an appropriate teacher preparation background. Service as a teacher in DCPS under a temporary certificate is limited to five years.

Certification is awarded in 77 areas of endorsement. Teachers can hold multiple endorsements. Endorsement areas are further identified by grade level being taught. The numbers of areas of endorsement by level are:

LEVEL	AREAS OF ENDORSEMENT
Pre-Kindergarten	1
Elementary (K-6)	10
K-12	6
Pre K-12	5
Junior High	6
Middle School	1
Sr. High and/or Career	

Development Center	13
Secondary (7-12)	25
Adult Education	5
Non-Teaching	5
Total	77

Teacher hiring in DCPS is based upon the candidate's ability to meet the certification requirements for the position being filled and by an additional review of academic qualifications, recommendations and a personal interview. Information concerning the competence of prospective teachers is limited to these sources of data.

Tenure is granted to DCPS teachers holding a standard certificate upon successful completion of a two year probationary teaching period. Teaching success is determined by receiving a satisfactory rating on the District's teacher appraisal process (TAP). Teachers are given five years in which to complete the requirements for tenure. The TAP is the major tool used for monitoring and rating teachers. It is relied upon for such major decisions as tenure, RIF procedures, salary increments and the identification of teachers needing assistance. However, the TAP is generally not viewed by either teachers or administrators as highly reliable. Both the 1984 Study of Teacher Incentives and the 1987 Career Ladder Review identified deficiencies in the design and use of the TAP.

It was the 1984 Study of Teacher Incentives, conducted as a part of a Congressional mandate to DCPS, that recommended a "new plan for selecting and inducting new teachers...to provide more rigorous screening and more intensive support." It was that recommendation that led to the development of the Intern-Mentor Program.

The report also recommended the use of tests of subject matter knowledge as a part of the DCPS "certification and/or selection practices." As a result of that recommendation, the Intern-Mentor Program, shortly after its initiation, contracted with the National Evaluation Services, Inc (NES) for the development of a series of subject area tests to be used as screening tests for the beginning teachers in the Intern-Mentor program. These tests, known as the Intern Content Knowledge Assessment Program, were designed to permit the intern and the mentor to determine areas of subject matter strength and weakness, and prescribe areas for further study and guidance.

These objective-referenced tests are based upon the DCPS competency based curriculum and were built using items and objectives originally developed by NES for use in other locations and tailored to the DCPS curriculum. (NES retains a proprietary interest in the tests.) The decision to follow this procedure,

rather than to use a nationally available test such as the National Teacher Exam (NTE), was based upon a desire to have a test equated as closely as possible to the DCPS curriculum within reasonable cost considerations.

By using previously developed test items the District has developed twelve content area tests and a writing exam for a cost to date of approximately $575,000 (other state test administrators indicate that typical costs for developing new subject matter tests for teachers average $150,000 per test.) Tests have been developed in the following twelve subject areas.

1. Early Childhood
2. Elementary Education
3. English
4. Social Studies
5. Mathematics
6. Science
7. English as a Second Language
8. Special Education
9. French
10. Spanish
11. German
12. Latin

During 1987-1988 NES is under contract to develop additional tests for the areas of Art, Music and Physical education and Health.

Developmental versions of the tests were administered twice within the Intern-Mentor Program. The operational version of the tests were administered in May and September 1987. In all four test administrations, only interns who voluntarily participated were tested. An analysis of the testing outcomes is included as appendix B.

In the three years the DCPS tests have been under development, the use of tests for hiring and/or certification of teachers has become more widespread in school districts and states across the nation. As of April 1987, 44 states required or were about to require passage of a written test before the award of full certification to new teachers: twenty-six states had a test available and eighteen more had tests under development or validation. Some individual school districts, including both Prince George's and Montgomery Counties, utilize their own tests in addition to state mandated tests.

Since initiating development of the NES tests, the issue of teacher testing has become more visible not only in the nation, but in the District as well. The press has raised the subject in its editorial pages as well as in its reporting and news analysis. Most recently, the Washington Lawyers Committee for

Civil Rights, in a report conducted in association with the Parents United, called for testing to ascertain the basic skills and subject matter knowledge of all teachers new to DCPS prior to their being hired.

In those states that have implemented teacher tests, experience to date has shown that minority applicants fail the first taking of the test at a disproportionately higher rate than Whites - in some cases the White passing rate is more than twice that for Blacks or Hispanics. However, there is some emerging experience to indicate that after retesting the passing rates become less disparate and fall within the boundaries of the "four-fifths rule" which the courts have used to hold that there is no adverse impact.

The implementation of any teacher testing program within DCPS must be considered within the context of the District's teacher demographic profile. As a recent DCPS strategy paper indicated, "there is now considerable evidence that the school system will face a teacher shortage during the next five years." Further, over the next decade, DCPS will replace a substantial percentage of its teaching corps. Accordingly, decisions and testing procedures adopted at this point in time are certain to have an important impact on both the District's efforts to recruit new teachers and its personnel profile over the next twenty to thirty years.

The recommendations which follow were formulated with all the complex considerations and the current status of the DCPS recruitment, certification, hiring, induction and tenure practices described above in mind. They represent a consensus judgement of the participants, based upon their considerable experience in these matters, as to the best procedures to follow in meeting the needs of DCPS students for quality instructional practice.

The recommendations also take into consideration the potential legal complications any program of teacher testing is liable to engender upon introduction to a school system. Appendix A contains an overview of these issues. It should be reviewed before proceeding to the findings and recommendations which follow in sections III and IV.

The findings and recommendations are grouped to present, first, those which relate to the development of an appropriate teacher testing policy in DCPS, given its current status and conditions. Second, a set of findings and recommendations are presented concerning test development actions which have been taken and which remain to be taken in order to appropriately implement the recommended policy.

III. TEACHER TESTING POLICY FINDINGS AND RECOMMENDATIONS

The sine qua non of a successful and defensible teacher testing program is the use of a test whose scores can be demonstrated to be valid for the purpose for which they are used. Therefore, the first step in developing a teacher testing program is to determine the purpose of testing teachers. There are many options. As a recent United States Department of Education Study stated:

> While virtually all the states have made some form of commitment to teacher testing, the diversity of philosophies and attitudes towards the issue are substantial. The states differ in terms of when they test prospective teachers, what their tests cover, the difficulty of their instruments and which tests are used. The issue of teacher testing covers not only whether teachers should be tested, but also how they should be tested and when. Virtually every aspect of teacher testing has been subject to considerable debate. (USDOE, 1987)

DCPS has been developing teacher tests to be used in an internship program for beginning teachers. The Board of Education has indicated an intent to develop a broader use of teacher testing. The authors of this report were asked to recommend an appropriate teacher testing policy for DCPS given the current status of its orientation to teacher testing and state of development of teacher testing materials.

In meeting our responsibility to make such a recommendation we interviewed experienced teachers, interns, principals, area superintendents, union leaders, central office executives, board members and interested citizens. We also conducted an extensive review of the currently developed tests and data concerning their use. We also surveyed our own preconceived notions as to the best use of teacher testing before we began our work on site. We have found a great deal of consistency among these three inputs, i.e. the DCPS perspective, our general orientation and the test development which has occurred to date.

In this section we first report our findings of DCPS's current status concerning teacher testing and related matters. We next offer seven recommendations for a teacher testing policy which we believe would best serve DCPS's responsibility, within its own unique local context, to provide quality instructional services to all its students.

FINDINGS

We have found in DCPS:

Broad-based support for subject matter knowledge tests, among the individuals and groups interviewed.

The participating consultants were issued a briefing paper and accompanying book of background materials before arrival in the District. From this information they developed a list of questions for the District staff and community members they were to interview. During their first session they refined these questions and agreed upon the information they were seeking.

During the day of November 18th the consultants spent about six hours interviewing approximately twenty individuals with varying responsibilities and interests in the DCPS. Some interviews were conducted in large groups. For some the consultants split into smaller groups. A schedule of interviews which identifies the interviewees is included as appendix D.

Those interviewed included central office executives, board members, community members, teachers, interns, union leaders, principals and area superintendents. Although the actual numbers of people interviewed were small they indicated they felt their views represented those held by other members of their role groups. There was agreement among all groups and virtually all individuals interviewed that testing teachers to determine that an adequate amount of subject matter knowledge was present was a reasonable and even desirable role for the District to perform.

Near consensus that the tests be used for initial certification and not for hiring, promotion or tenure decisions.

DCPS is relatively unique in the sense that it is both a school district and a state level agency as well. Only Hawaii and the U.S. territories share this dual role. In this dual role the District serves as both the teacher certifying agency and the hiring agency. These functions are performed in separate divisions. In other states the state department of education serves as the certifying agency and the local education agencies serve as the employing agency. Forty four states are currently using or considering the use of some type of standardized test for certification purposes.

Our discussions with DCPS interviewees indicated strong consensus that subject matter knowledge testing be used for certification purposes and not for employment related decisions such as hiring, promotion or tenure. There seemed to be a clear sense that these were separate functions in this district and that testing was more appropriately a function of the public protection role of the state department side of DCPS than the

employment side which seeks to maximize hiring decisions. As much as anything else, this sentiment seemed to reflect an awareness of the limits of testing as a selection device, especially in as complex an endeavor as teaching.

Substantial sentiment in favor of a short grace period during which a candidate would be allowed to teach, not withstanding that the candidate had not yet passed the certification test.

Consistent with this awareness and sometimes even suspicion of the ability of testing to predict job capability, the interviewees favored providing certification candidates an opportunity to bring up their performance before positively barring them from the classroom. There was a strong sense that the absence of significant evidence of predictive validity relative to standardized testing and teaching ability combined with the well documented observation that minorities typically score less well than others on standardized measures, argued for tempering the initial impact of testing on those seeking to become teachers. There was also a recognition that this argument was less persuasive when dealing with tests of subject matter knowledge than when dealing with other less specific areas such as pedagogy.

From this tension there emerged the notion of a grace period during which a temporary certificate could be issued for those candidates actually hired by the district. Just as such a certificate now exists to permit those who fail to present all the requirements for certification an opportunity to demonstrate their competence and obtain the required coursework, etc., so would the issuing of a temporary certificate for those who did not pass the certification exam on the first attempt permit them to obtain the necessary assistance to demonstrate their subject matter knowledge.

There was also the sense that during this period of temporary certification, the District should have the responsibility to both closely supervise the teacher and offer assistance in meeting whatever deficit the test failure implied. Of course, any utilization of this temporary certificate option would have to be utilized with discretion and judgement relative to the relationship between the assignment and the test performance problems. There was no agreement on the length of such a period but the range of one to two years seemed to be about what people had in mind.

Support to require testing of all persons seeking initial DCPS certification, including persons previously certified in other jurisdictions and also including persons voluntarily seeking additional endorsements.

There was little debate that the certification testing requirement should be imposed in a uniform manner. If "beginning" teachers should be required to demonstrate their subject knowledge, so too should experienced teachers new to the District or "voluntarily " seeking additional subject area endorsements.

While it was recognized that states routinely offered "reciprocity" of certification, it was clear that the interviewees felt reciprocity should be limited to the other certification criteria, and not extend to the requirement to demonstrate subject matter knowledge. The only exception to universally using testing for certification was to those situations where a teacher was seeking an additional endorsement because of a district imposed reason. Clearly, some guidelines must be worked out in these areas to protect students from teachers who do not have the necessary subject matter knowledge.

A view of test passage as a necessary, but not sufficient, requirement for certification which should be based on additional factors as well, including, e.g., recommendations, GPA, coursework, successful student teaching.

The use of subject matter testing was seen universally as an addition to the current certification requirements, rather than a replacement. Subject matter knowledge, as demonstrated through test performance, is only one element of the make-up of a quality teacher. The other criteria for certification - graduation from an approved training program, minimum GPA, minimum major area GPA, and successful student teaching experience, are all important measures of other qualities sought.

Since the goal of testing is to assure subject matter competence, not deny employment, it is hoped that all candidates pass the test. Therefore the use of other measures remains as important as ever. This is especially the case when dealing with minority candidates who may have a difficult time initially with the test. The other measures provide additional information to the temporary certification and employment decision. Support for the use of multiple measures spanned all groups interviewed. Testing was universally seen as a necessary but not sufficient condition for certification.

A perceived need, even with the addition of testing to the certification process, for improved procedures to support other decisions such as personnel selection and tenure.

Despite the universal feeling that testing should be restricted to certification or licensure purposes, there was also nearly universal concern that other employment related areas of the District's operations required attention. There are problems in educational organizations, generally, related to the absence

of an objective and fair definition and rating device to identify effective teaching. DCPS is no exception.

Though some respondents would equate the promotion, tenure and pay decision directly to student performance, others felt the current system is adequate, if at times unfair. Generally, however, most interviewees felt the District should revise its teacher appraisal system and strive to make its hiring practices more objective. Related to this issue was the strong belief that administrators eschewed the "tough decision" and rarely denied tenure or attempted to dismiss the tenured teacher who had "lost it". There was clear dissatisfaction with this situation in all quarters even in the union, which, nevertheless, felt its contractual provisions for due process were necessary in the face of equally valid concerns with fairness and retributive behaviors of some administrators. Clearly this is an area for further work.

A lack of clarity with regard to the purpose for which the writing test was developed and how the DCPS intends to use it.

In interviews with the Intern-Mentor personnel as well as with the NES representative the purpose of the writing test was not clearly identified. We remain unclear as to the central purpose for which that test was developed. This is particularly disturbing given the often very subjective, time consuming and expensive scoring of such exams.

We recognize that school systems are vulnerable to extreme criticism over the writing abilities of their personnel. It is easy to sensationalize such problems, and they do, indeed, exist. However, we are not clear about the intended purpose of this test in their resolution, nor does DCPS appear to be.

RECOMMENDATIONS:

As we stated in the introduction to this section, the most important element of a testing program is a test whose scores are able to be demonstrated to be valid for the purpose or inference being made of them. This, therefore, requires as a first step in the development of a testing program, a clear statement of the purpose of the program. We have reported our findings concerning the orientation of DCPS involved interviewees towards teacher testing. We concur that the testing purposes described by the respondents are the best ones for DCPS to hold for a variety of reasons, including its current teacher recruitment needs, legal considerations involved in test development and use, its responsibility to serve its students and the current status of DCPS test development efforts.

We, therefore, recommend DCPS adopt a set of interrelated teacher testing policies which clearly direct the purposes of its teacher testing practices to its licensure function and responsibility.

Specifically, we recommend:

Subject matter knowledge testing of teachers, aligned, to the extent possible, with existing certification/endorsement areas, be used in DCPS for purposes of certification or licensure.

We concur with the consensus in DCPS that subject matter knowledge testing of teachers should be used for purposes of licensure or certification, rather than any other. We also believe that if a jurisdiction wishes to use the testing of teachers as a public safeguard device then subject matter testing is the best choice.

Subject matter knowledge is measurable in reasonably objective ways, most agree it is a necessary condition for effective teaching, and its a higher order demonstration of knowledge than basic skills exams. General knowledge exams are not job related and therefore are not legally very defensible. And, professional knowledge is really better demonstrated and remediated in on-the-job situations than through testing.

The use of such testing for certification rather than employment purposes is both legally and psychometrically an easier and more defensible route to take. This concept is expanded further in Appendix A which should be reviewed now if it hasn't yet been read.

The major issue DCPS faces in using subject matter testing is determining the specificity of the tests utilized. There are 77 certification areas. Clearly, the District cannot afford to develop 77 subject tests. However, neither is it fair to include, for example, knowledge of physics for candidates for certification as biology teachers - as the current tests seem to do. We have not been able to determine how the current testing areas were identified, however, we urge a careful review of the certification areas, the expected areas of vacancy, and the availability of tests from other jurisdictions before further resources are devoted to test development. The next section also addresses this issue.

In order to be granted a standard teaching certificate in DCPS, a candidate must achieve a specified score on the certification subject-matter test.

Testing for certification purposes is not intended to rank candidates as does testing to select the best employee. Rather,

it is intended to determine that a candidate for a licence to teach in DCPS can demonstrate that they possess at least the minimum subject matter knowledge necessary to teach effectively in the District's schools. While some jurisdictions might choose to utilize a compensatory model, wherein some other factors such as grades or student teacher ratings would serve to add points to a candidates score, we rejected recommending such an approach. There are no satisfactory substitutes for deficiency in subject matter knowledge testing, given adequate opportunity for retaking the test.

Therefore DCPS will have to determine cut scores that reflect this criterion. The cut score should have some theoretical and empirical basis to it. Additional field tests will have to be conducted to develop an adequate data base to make such decisions. Section IV discusses these procedures in greater detail.

The certification test requirement apply to all persons seeking initial certification in DCPS, including persons previously certified elsewhere and teachers seeking additional endorsements

We concur with the consensus we have reported and the reasoning supporting it. Previous certification elsewhere is no substitute for the validated demonstration of subject matter knowledge testing provides. Neither does previous certification in a different field. We support consistency in the use of tests when they are an available and cost effective option.

Certification candidates who have not achieved the specified standard on the certification test, either because of failure or lack of opportunity to be tested, be granted a time limited exemption from the certification test requirement during which time the candidate, if otherwise qualified, would be granted a temporary certificate and permitted, if hired, to teach in DCPS.

DCPS does not "own" the subject matter tests currently developed. NES retains a proprietary interest and will charge for test administrations. Therefore, it will not be feasible to offer test taking opportunities more than two or, at best, three times yearly. This fact, plus our knowledge of the first time passing rates of minorities, verses their eventual passing rates, virtually mandates this recommendation. To do otherwise would unrealistically inhibit DCPS's teacher recruitment efforts in what is already showing signs of becoming a difficult situation.

During a period of exemption DCPS should use its auspices to assist candidates to pass the test when it is taken. As candidates teach, especially with proper supervision, they will become familiar with the subject matter requirements of the

District's curriculum. They should also be given opportunity by the District to hone their test taking skills. DCPS should not be expected take on the teacher preparation institution's role. However, it can offer test taking seminars and other like forms of assistance. It is in the District's best interest to utilize a rigorous testing program and assist as many candidates as possible to pass it. DCPS's goal in testing teachers is not to deny candidates a teaching role. It is to provide the best possible teachers to its students.

Candidates teaching without having taken the certification test be required to take the test at the first available administration after their employment.

None of our recommendations for exemptions are meant to diminish the impact and quality of the District's future teacher testing program. They are offered to strike the proper balance among the various factors which must be faced in meeting DCPS's responsibility to provide quality instruction.

We recommend the testing program be implemented and administered with a fair and firm hand. All new candidates for certification must take the tests as soon as possible after their candidacy is initiated. There should be no excuse for not doing so and failure to do so should be met with consequences concernig future employment.

DCPS develop appropriate modifications to teacher contract forms to be used for candidates teaching without having passed the test, in order to take into consideration remedial, administrative and legal ramifications of the arrangement.

This recommendation is intended to avoid problems of "due process" down the road. The teacher testing policy in DCPS must be clearly stated, disseminated and reflected in all relevant areas of the district's operations. Teaching contracts should provide clear statements of the conditions of employment with respect to test taking, timeliness, release of test scores or evaluations for support and remediation purposes and a clear understanding of the time limited quality of any exemptions and the consequences related to future employment of exceeding those limits.

DCPS undertake an analysis of its current appraisal processes used in hiring, promotion and tenure decisions and develop an overall evaluation system that leads to either further professional development and/or non-renewal and discharge decisions.

The pervasiveness of concern expressed regarding the employee selection and evaluation practices of the District are striking. While most acknowledged that "politics" plays less of a

role than in the past, issues of favoritism still pervade the process. While it is virtually impossible to render any large organization free of these charges, there is objective reason to suggest that DCPS can make further progress in this realm.

Two objective outside reviews have identified the TAP system as being flawed in both design and use. It was beyond the scope of our charge to review this system, but, the level of concern we did hear expressed motivate this recommendation to address the issues in selection and promotion which go beyond testing.

In doing so the District likely will be able to reduce its reliance on testing as a screening device. There is reason to believe that hiring decisions in education can become increasingly reliable as they become more objective and involve more individuals from more diverse settings in the district. Addressing this issue can, therefore, deal with both the testing and the teacher professionalism issue as well.

The certification testing program include the NES tests currently in development. For areas in which no test is being developed, options which should be considered include: 1. no test be developed, particularly in low incidence fields; 2. validation of existing tests (e.g. NTE); and 3. development of additional tests through the NES process.

While it is understandable that DCPS should want to have teacher tests which are tailored to their curriculum, this is, potentially, an expensive undertaking. As we have noted, there are 77 certification areas. Fairness demands some relationship between these areas and the test coverage. Developing 77 tests is not a cost effective undertaking, especially in a district that only can be expected to hire about four to five thousand teachers over the likely life of a set of tests.

Therefore, we recommend strongly, that for areas in which few teachers will be hired, no test be developed. In these areas the payoff from the development of a more complex and objective hiring process can be relied upon. In other areas, DCPS should first review extant tests and attempt to validate them for use in the District. With forty-four states testing teachers, more and more tests will be available. Local validation, per se, is a relatively inexpensive endeavor.

The development of additional custom made tests should be approached only as a last resort. We simply do not expect the numbers to merit the investment. However, please be clear, we have not suggested the substitution of non-subject matter tests. We do not believe the state-of-the-art in their development yet renders them a reliable tool, especially in a jurisdiction which draws from so many others in developing its candidate pool.

The writing test not be a certification requirement, but be required of all newly-hired employees and used for diagnostic-prescriptive purposes.

As we have stated, the purpose behind the development of the writing test is unclear. Further, its application has some inherently subjective elements. We are not comfortable seeing it used in a manner that could be a bar to employment. However, it does provide some useful information to an employing agency.

Therefore, we recommend that it be utilized, where necessary, to support staff development for newly hired DCPS professional employees. This will require the District to develop some programs to remediate deficiencies, however that would be a worthwhile endeavor. In this manner the development of the writing test could have some significant payoff for DCPS.

IV. TEST DEVELOPMENT PROCEDURES FINDINGS AND RECOMMENDATIONS

The preceding section stressed the importance of having a clear and consistent purpose for testing teachers as a basis for the development of tests which will be able to produce scores which are valid for their use. At the present time, such a purpose has not yet been formally articulated by DCPS, although a strong consensus as to the best use of teacher testing appears to exist. A set of policies consistent with that consensus and which would provide the basis for a sufficiently specific statement of purpose to permit test development to be completed satisfactorily was recommended.

In this section we report our findings concerning test development practices and procedures to date. In general, we find that the twelve subject matter tests developed by NES require additional development work to permit them to satisfactorily serve the purpose recommended. We go on to suggest a series of steps which will complete the test development process in a manner consistent with the state-of-the-art in teacher testing and permit the use of the extant tests for the purpose described in section III.

FINDINGS:

Concerning the test development process, we have found:

<u>More documentation of the test development process and activities carried out by NES is desirable.</u>

The gravity of the decisions to be made using scores from these tests required appropriate documentation of the process of test development. Interviewees made reference to various pieces of information from NES regarding test development, but there is a need for greater organization and accessibility of these data.

A manual covering test development activities in detail and providing data collected from those activities has not been provided. There are individual elements of information, but these data have not been assembled and edited to provide a coherent map of the test development process. The <u>Standards for Educational and Psychological Testing</u> provide overall structure and background for the development of this material.

<u>The test development strategy involved the collection of data from beginning teachers in a low-key and non-threatening manner.</u>

The intern-mentor program provided the framework for the collection of field data for the testing program. Interns were invited and encouraged, but not required, to take the tests. Results were used initially by mentors and interns to review

strengths and weaknesses, and to provide necessary activities to strengthen weaker areas. Interns seemed to feel that they had gotten a lot of help from mentors in improving in these areas. It was indicated that results from more recent testings have not been disseminated as quickly, and that the feedback has been desired by the test takers.

These field test results were not used in making hiring or placement decisions of interns, a fact which aided in their acceptance of the testing program. The fact that not all interns were tested does reduce the amount of data available, and causes questions about representativeness of the available data. No data seems to have been collected regarding retesting of interns and their classroom performance.

Further test development activities with appropriate documentation are needed.

In order to create a teacher testing program in which test results can be put to valid use in the licensing of teacher candidates, the instruments must have strong grounding in the indices of test quality. With an instrument designed to have content validity, it is important that the activities used in establishing and revising the test objectives be carried out in a manner which assures the validity of the process. It is equally important that the same standard apply in validating test items in terms of these objectives.

Instructions given to teacher-raters in the objective preparation and in the item validation conferences, insofar as we were able to ascertain them, do not appear to have been sufficiently specific. While the objectives were reviewed, and teachers did rate each of them in terms of use, teaching, time spent and the degree to which each was essential, there does not seem to have been a systematic procedure for tying these objectives to the DCPS curriculum. Teachers were asked to recommend additions and deletions in the set of objectives, but this was not an activity that focussed specifically on a comparison of the complete set of the objectives and the DCPS curriculum.

In the item validation conferences, all items, according to interviewees, were reviewed for appropriateness for the beginning teacher, the extent to which the knowledge was needed in order to teach, and other relevant criteria. However, it does not appear that reviewers had an opportunity to review the full set of items for balance, emphasis, and overall curricular relevance. Also, the validation was seen as a mechanical rather than intellectual process by at least some of the teachers who participated.

Items received a review of their content for apparent bias. There were two reviewers for each test, chosen from specialists

in related subject areas. It is not clear that any of the reviewers had specific training in detecting item bias.

The question of the fit of the tests to the curriculum was raised in the interviews with interns, with mixed results. Some tests were described as relating well to the curriculum. Others, were not. The sub-area scores were found to be useful in identifying content areas in which the teacher needed further review. The science test was criticized because it covered all physical and biological science areas, rather than providing separate tests for physics, biology, and chemistry. Language tests were described as not relating modern romance languages to the cultures of African and West Indian countries that speak the languages.

<u>There has been insufficient technical monitoring of the development of the tests.</u>

Supervision of the initial phases of test development has taken place at the level of the Superintendent's staff. This has allowed for a low-key field data collection effort and encouraged broad support, while emphasizing the commitment of the Superintendent to the activity. To complete the necessary test development, more technically-oriented monitoring of the test development effort is needed. The framework set in the past two years should ease overall acceptance of the testing program.

<u>The purpose of the tests has not been stated with sufficient clarity.</u>

The conceptions held by interviewees regarding the purpose(s) of the tests varied considerably. There was a general sense that the tests would be used for initial certification. One administrator found the tests a practical way of determining whether a person had the potential for teaching. Interns and mentors tended to view the tests as diagnostic-prescriptive instruments. The community representatives felt that tests should prevent entry of teachers not meeting standards.

There seemed to be uncertainty about whether higher levels of a test score might be used for distinguishing between candidates for hiring, or whether the scores might have a role in subsequent career advancement.

DCPS and NES must fully articulate the test purpose and make its intent clear to all involved parties. This is not to suggest that the test might not have some usefulness outside of its main intent. Results may certainly be of use to interns and mentors in further developing the competencies of interns, even if diagnosis is not the purpose of the test. However, a clear and

consistent purpose for teacher testing must be articulated and observed.

An insufficient amount of data has been collected for use in standard setting.

The numbers of beginning teachers who have taken the tests is quite small for most subject areas. The setting of standards, or "cut off" points requires larger field testing samples.

The existing data set is based on volunteers who may or may not have achieved scores similar to that which those who were not tested would have received. While interns were encouraged to take the tests, and many did, there were many who did not. This raises the issue of the representativeness of the test takers to the entire set of interns and potential applicants. It also results in a smaller number of cases with which to make standard setting decisions.

When sufficient data have been collected, standard setting reflecting the validation and the field testing data can be undertaken using multi-step procedures.

Participation of administrative and policy personnel has been limited.

Interviews with policy and administrative personnel demonstrated modest participation in the test development. This group was familiar with the content-validity approach used, but was unfamiliar with the procedures used by NES in test development. As stated earlier, there seemed to be a lack of clarity of the purpose of the tests.

This group expressed opinions related to test use policy, such as the belief that test should be only a part of the requirement for teaching and that testing should be done early in the process of candidate review. There seemed to be a shared belief in the importance and usefulness of a testing program. Some expressed a desire for the development of tests of writing and speaking skills. Greater consistency should be achieved as the testing program moves closer to implementation.

There is a concern that any analyses of test results should take into consideration the impact on minorities, and the realities of teacher supply and demand.

Several interviewees noted a concern for the fact that the majority of DCPS teachers and teacher candidates are Black, and that minority status tends to be related to lower performance on standardized tests. There was a concern that qualified minorities continue to join the teaching force, and that the test not negatively impact their movement into regular teaching

positions. There was a sense expressed that tests and the process of developing them were fair, but the concern remained that minority teachers might be adversely affected. These concerns cut across several sets of interviewees.

The realities of supply and demand were raised by several interviewees. They noted that the school district is surrounded by suburban jurisdictions with better-paying school districts and with working conditions that in terms of physical plants and the demands of personal involvement in the job of teaching may be perceived as more positive. These realities are already pulling away good teachers from DCPS, and a concern was expressed that this trend might be accelerated with the introduction of beginning teacher testing. The fact that most other jurisdictions either are or soon will be using tests for certification should lessen this problem. However, it was noted that if DCPS uses its own tests rather than a broadly-used set of tests such as the NTE, this might deter candidates from applying. Several interviewees including the community representatives supported the use of the NTE, or the option of the custom-made or NTE tests.

RECOMMENDATIONS:

It has been recommended the DCPS teacher testing program become a component of its teacher licensing (certification) functions. As we have described above, for a variety of reasons, the tests which have been developed for use in the District's Intern-Mentor Program, do not yet meet the standards for use as a licensure examination. It is our expectation the deficiencies identified can be remedied by the collection and codification of additional data on the test development process used by NES and by further test development procedures to validate the existing tests for use in a certification function.

We, therefore, recommend DCPS engage in a series of test development actions which are clearly directed to developing the validity evidence necessary to support the use of its subject matter tests for licensure purposes.

Specifically, we recommend:

<u>The Superintendent request NES to immediately inventory and submit all available documentation to DCPS for review.</u>

For any test there should be adequate documentation of the test development process. This is particularly important in high stakes tests such as the proposed licensure test. The <u>Standards</u> (AERA-APA-NCME, 1985) have a chapter on "Test Publication: Technical Manual's and User's Guides" which contains 11 standards. The index to the <u>Standards</u> references other standards

that pertain to publisher's materials. The background section of the chapter makes the following relevant points:

> "Publishers should provide enough information for a qualified user or a reviewer of a test to evaluate the appropriateness and technical adequacy of the test.
>
> Even when a test (or test battery) is developed for use within a single organization, a brief manual will be useful." (p.35).

There has not been total agreement either about the degree of detail that should be in a manual or about the kinds of statements that a publisher should be able to document. In the Alabama lawsuit on their teacher licensure examination (developed by NES) the documentation issue received considerable attention. Plaintiffs' experts argued for the necessity for very complete documentation and NES could not, in fact, supply all the documentation they requested. Although the Defendant's experts were not opposed to documentation, they felt that a rule of reason should apply and that many of the plaintiffs' experts requests were not reasonable. Nevertheless, because critics will want to audit the test construction process for high stakes tests such as licensure tests the publishers would be wise to be particularly diligent in the accuracy and thoroughness of their documentation.

The documentation we have seen thus far has been sufficient. For example the Lawrence Johnson report does not inform us regarding the return rate of the objectives rating form. The actual form and the accompanying instructions are not presented in the report. The Interim Content Knowledge Assessment Program--Final Report does not include the specific information that is required. Further, the DCPS does not know whether the documentation files to back up the data in the report are available.

Those who contract with publishers have a responsibility to request sufficient documentation. Hence, our first recommendation is that DCPS receive all previous test development documentation for review.

DCPS with outside consultation, conduct further technical and legal reviews of this documentation to guide future development.

Because the Standards suggest documentation, and because critics will demand to see such documentation, it is incumbent on the DCPS to review the existing documentation to determine its adequacy and to use such documentation to guide future development. If documentation for some essential test development process is lacking the DCPS may wish to request that process be redone and documented to assure adequate quality of

the test. Because there is always a chance for litigation, the DCPS legal counsel should be involved in this review. Suppose, for example, the results of a "bias study" were reported, but there was no documentation to support the methodology or the findings of the study. It may be wise to redo such a study so that documentation is available.

The suggestion for outside consultation in the recommendation is based on our belief that any organization, no matter how well staffed, should have such consultation for high stakes tests. This is certainly a fairly common operating procedure, and we believe a wise one. Because many states have been involved in building and administering teacher licensure tests there exist many experienced experts who could provide useful insights to the DCPS staff as exemplified by the convening activity.

Specifically, we suggest that a team of at least three consultants be asked to meet with members of the DCPS technical and legal staffs for a 2 to 3 day meeting to go over all existing documentation. This outside group of consultants should include at least two measurement experts and one attorney who have some experience in licensure tests. We suggest that the team look for documentation such as that listed below. First, it should be determined whether documentation exists. Second, it should be determined whether the quality of the procedures which have been documented are sufficiently high. We stress that the list below is not exhaustive--it is presented to give readers a general idea of the documentation we would hope to see:

1. Is there adequate documentation of all contractual agreements between NES and DCPS?

2. Is there adequate documentation that all aspects of the agreements have been carried out?

3. What documentation exists regarding how the various DCPS personnel were selected to participate in various portions of the test construction/validation process?

4. Is there documentation regarding various characteristics of these DCPS personnel (e.g., sex, ethnicity, age, experience, education, etc.)?

5. Are all existing data from the various committees kept as original copy or do the data only exist on data tape? That is, do the raw, disaggregated data exist or is only summary data available?

6. Is there existing documentation regarding the specific instructions given the various committees?

Once the committee of inside and outside consultants have obtained answers to such questions, they need to decide what future steps should be undertaken prior to using the NES tests.

Future development should be undertaken and accompanied by written documentation to ensure all aspects of test quality.

Because we are not knowledgeable about all NES has done and how much of it they can adequately document, we cannot specify all the details of future development activities. Nevertheless, we are aware of several steps that we believe should be taken. We talk about some of these in more detail under recommendations 5, 6, and 7 in this section. Those tasks relate to further validation, field testing, and standard setting.

In addition, we would recommend that NES provide statistical bias data for the questions based on the Georgia administration and that these data be reviewed. Specifically, we would suggest that for all items being considered for the DCPS test, there be an empirical bias study using one of the more commonly accepted approaches (e.g., using ETS's Differential Item Performance procedure) and based on as much data as exists from Georgia on those items. The results of that study should be considered by the DCPS technical staff, by outside technical consultants, and by a committee of DCPS teachers to determine whether the items should be included on the DCPS test.

Further, the reliability of the test and the reliability of the decision need to be analyzed. Of course the reliability of the decision can not be analyzed until the cut score is set, but we mention it as one more aspect of future test development. Because the test is to be used for licensure decisions, it is necessary to estimate the consistency of the decisions. As the Standards point out:

> "Estimates of the consistency of decisions are needed whenever decision rules assign people to categories according to specified test score intervals. An estimate of the standard error of measurement at the cut score is helpful" (AETA, APA, NCME, 1985, p. 20).

Although there is no single formula specified in the Standards it is probably fair to say that most measurement specialists would feel that the Subkoviak, Huynh and Marshall procedures are all acceptable.

Another approach we recommend would be to estimate the reliability of the domain score estimates--consistency across parallel or randomly parallel test forms. The traditional K-R 20 is commonly used if one assumes parallel tests. As Traub (1986) points out, although such an estimate is not required by the Standards for licensure tests, it does provide useful

information. It does not replace one of the other estimates discussed earlier.

Although DCPS and the contractor developed the tests using a statement of purpose which appears consistent with the use of the test for initial certification, a more precise statement of the purpose should be used in collection of additional validity evidence.

A test developed for one purpose will not necessarily be appropriate for another purpose. Test validation is the process of accumulating evidence to support particular inferences, and thus one must be aware of what inference is hoped to be made from the test scores at the time the test is being developed. It is clear that NES knew they were "to provide examinations for beginning teachers in the District of Columbia public schools" (NES, 1987, p. 1). Further, it was clear that the tests were to measure content knowledge in the specific teaching fields. However, it was also suggested that "these tests may help the intern and the mentor teacher to evaluate subject-matter strengths and weaknesses and to determine areas for further study or concentrated guidance" (p. 1). The Content Inventory Analysis Final Report stated that "The tests presently under consideration will serve as an entry level measure of each teacher's mastery of the curriculum and content for which he or she has been trained" (Lawrence Johnson & Associates, 1985, p. 1). Individuals we interviewed did not always use the exact same words when discussing the purpose of the test. While we do not view this with any great alarm, it is clear that critics of the test and the validation processes used in the test development will look closely at the statements of purpose used by the test developers. For example, in the Alabama lawsuit the Plaintiffs' expert witnesses made much of an Alabama Board statement that their test was "to measure the specific competencies which are considered necessary to successfully teach." Those expert witnesses tried to suggest that the phrase "to successfully teach" implied that the test should have criterion-related validity and that successful is a matter of degree that can be measured along a continuum among those who are qualified. While we believe a more reasonable interpretation of "successful" as meaning above a minimum cut score (which is what one wants in a licensure examination), and that in fact "considered necessary" correctly suggests that the decision is a professional judgment and that competency while necessary is not sufficient to "successfully teach," the debate in the Alabama court suggests that it would be wise to be as specific as possible in labeling the test as a licensure or certification examination designed to measure whether individuals have the necessary subject matter knowledge and that such a test is for the purpose of protecting the public from incompetents, not for predicting success as a teacher. This understanding of the purpose should be clear to all who participate in further validation.

Validation of the existing test continue and include additional input from teachers, university specialists, measurement specialists.

In its test development process, NES followed the basic steps of producing valid licensure/certification tests. The establishment of content validity is fundamental to the development of licensure tests. Content validity is only established in test construction (Cronbach, 1980). The five stages in establishing content validity of a licensure test were set forth by Mehrens (1987) and include:

1. developing an original list of competencies,

2. doing some type of job analysis survey,

3. specifying (and validating) the domains (objectives) for the test,

4. writing and validating the items, and

5. obtaining an overall judgment of the content validity of the test.

Subsets of these five stages are bias reviews and the establishment of a committee criterion for making decisions about item quality and test coverage. Although NES based its validity work on the basic premise of these five stages, additional validation procedures must be implemented to support the content validity of the tests for the purpose recommended herein.

The DCPS process is unique in that most groups establishing certification tests either (1) select a pre-developed test and validate it for their purpose, or (2) start from scratch in the development of their own tests. Rarely does one attempt to adopt/modify an existing test. What follows is a discussion of the enhancement of content validation procedures as they relate to the DCPS present and future test development efforts. The emphasis in this discussion focuses on the current validity procedures and a recommended a one-day revalidation process to validate the current DCPS tests for the purpose recommended in a manner consistent with current state-of-the-art validation procedures.

There are three groups of tests that have been or will be established by the DCPS. The first group includes those that have been validated "as is" or with the addition of new items as well as those tests that were developed solely for the District of Columbia. Secondly, there are those tests that have been developed elsewhere but will be validated in DCPS for their use in the future. The third group consists of the DCPS tests that

will be developed in the future. Recommendations for each of these three groups follow.

Current DCPS Tests

Our concerns about the tests that have been developed by DCPS include (1) the lack of a clear purpose for the test prior to validation (addressed in Recommendation 3), (2) the small numbers of individuals on the validation advisory committees, (3) the use of a 51% criterion for decision making, and (4) the procedure used to calculate the results of the job analysis survey.

In the review of the job analysis information, another problem was encountered concerning the procedures for calculating the results of this survey. Apparently data were gathered on four variables:

a. objective taught or not,

b. objective utilized or not,

c. time, and

d. essentiality.

If teachers responded positively to A or B, their responses for C and D were tallied. However, many did not respond positively to A or B. We do not know what algorithm was used to combine C and D. While this often does not matter in terms of the results, a few licensure experts would argue that the percentages obtained in A and/or B should be included in the algorithm; and many would argue that D should receive greater weight than C. More attention needs to be given to this matter in future validations.

To rectify the job analysis concern and the concerns in other areas, we recommend that a revalidation occur based upon a clear statement of purpose. A new advisory committee for each test area should be formulated. These committees should consist of a minimum of 15 to 20 members each, where possible. (Areas such as Latin and Spanish are clearly exceptions to the rule.) The composition of a committee should consist of approximately 11 to 16 teachers, two district curriculum specialists, and, possibly, one or two local university content experts. To the degree feasible, attempts should be made to assure that each individual school faculty in the district is represented on at least one advisory committee for the revalidation and development processes.

A new decision criterion should be established and used in the revalidation process. The 51% majority decision employed by

NES allows for too much concerv over the validity of the decisions made. A two-thirds or preferably three-fourths majority criterion is most frequently selected by states in validating NTE's. ETS uses a 60% majority decision point.

To shed light on these issues further, we recommend that DCPS use a more current procedure for revalidating the certification exams. The job relevance and bias issues should be the focal points of the revalidation effort. An example of a revalidation procedure and the accompanying judgments forms (which were developed by Instructional Objectives Exchange) are provided in Appendix E.

A procedure of this type usually takes less than one day for an area exam. The panel or an advisory committee of 15 to 20 teacher/curriculum specialists should be brought together. Each new committee should receive an orientation regarding the purpose of a licensure test and the differential validity requirements of a licensure test and an employment test (see the Standards, 1985).

If two forms of a test are being validated, the procedure should entail (1) item-by-item judgments for the first form, (2) total test judgment for the first form, item-by-item judgments for the second form, total test judgment for the second form,a nd form-equivalence judgment.

Once the judgments are made, the consultants or contracting agency in charge of the project should calculate the data and provide summary data to DCPS. These data should include numbers of panelists, numbers of items, % of items at or below the 50% job relevance, % of items at or above the 75% job relevance, total test judgment range, job relevance average per item index, majority index, and bias-average per item index. DCPS should review the data and present it to the Board of Education. If a consultant committee is established, the consultants should provide DCPS with recommendations for the Board. If DCPS cannot recommend a test, then the recommendation to the Board could be to not use the specific test and validate another existing test or to not use the specific test and redevelop a test.

The revalidation procedure suggested is both efficient and cost effective. It can be accomplished for a cost of $10,000-$12,000 per test if contracted and possibly less if handled by DCPS consultants.

Future Validation of Previously Developed Tests

For future validation of the subject area tests developed elsewhere, we recommend the same validation process suggested for the revalidation of the current DCPS tests. At this stage, the development of new items and a job analysis survey are not

necessary. If the test validate, they can be administered; and the standards can be set after the first "live" administration of the test.

The Development of New DCPS Tests

The development of new DCPS would entail the five stages in establishing content validity as prescribed by Mehrens (1987).

An advisory committee of 16-20 members (the composition of which was described earlier) should convene to establish a set of domains/objectives for the area examination. If a contractor is used, the contractor should provide the committee with the domains/objectives of the NTE's and NES tests and various other state developed licensure tests, as well as a list of the domains/objectives that have been included in the literature.

A set of domains/objectives should be identified for a job analysis survey, and a job analysis survey should be conducted.

Upon completion of the job analysis survey, the advisory committee of 15 to 20 participants should be reconvened to review the job analysis survey results. The committee members should be given the total data, not just the weighted averages. This information should be used to select the domain/objectives to be tested and make initial judgments of the weightings of the objectives for the test based upon importance of the content knowledge. University staff can provide external content knowledge input into the process. These initial objective weightings will give the committee some feel as to what items must be developed. Once the objectives are agreed upon, the committee should develop a set of test item development specifications which would contain the skills to be assessed, appropriate vocabulary ranges, etc. These are the blueprints that are necessary to guide item development.

Once the objectives are given a preliminary weight and the item specifications are developed, the test items for the objectives should be developed. Additional items should be developed for item replacement purposes, and the preliminary forms of the tests should be constructed. Two external reviews should be conducted on each test developed. These reviews should consist of a technical item review by a measurement specialist and a bias review by two or three bias specialists (using a minimum of one white and one to two Black bias reviewers). An external subject matter review by a content specialist would not be necessary if local university content specialists serve on the committee. The 15 to 20 person advisory committees should be reassembled for the item review. During the item review, the external measurement specialists and one of the bias reviewers should be present. The measurement specialist should provide approximately a 45-minute training session on detecting technical

problems in items, using example items containing problems. The regular NES process can be used in making judgments about each item with the addition of a "range of difficulty" category. Also, a comment column should be provided for committee members to note technical and content problems found in the items. The committee should make judgments as to whether to "keep," "revise," or "delete" each item. The committee members should, without conferring with other committee members, read and answer each item. An answer key should be provided separately for more accurate decisions about the items. (Individuals tend to assume more knowledge about an item if the correct answer is keyed on the item.) The items should be reviewed orally by the group to make the final item decisions. During the oral discussion, the measurement and content specialists and a bias expert should provide their particular input as each item is discussed. The committee established criterion should be used for final decisions about items. If an insufficient number of items remain, new items must be developed, using the objectives and their weightings and the test item specifications.

Further field testing should be conducted in areas with sufficient numbers of applicants to have meaningful data for further refinement of tests and possibly for standard setting.

The development of a multiple-choice test is a time-consuming process. Although a sound approach to assessment, this approach is not cost effective for low incidence subject areas. Alternative approaches to testing might include: (1) performance assessment using at least two trained judges or (2) validation of pre-existing test where sufficient numbers of individuals nationally have taken the tests. Otherwise, item quality through the empirical validation process is impossible. Unfortunately, making judgments on the quality of the items and standard setting without item data is further complicated by the small number of teachers in these areas available to make judgments.

For the tests that have sufficient numbers of examinees and committee members, all newly developed items should be field tested. The advisory validation committees should reconvene to make final decisions on the test items based upon this empirical data.

Committee members should be provided with the percent of correct data for the total group and for subgroups, including race, sex, and upper and lower quartiles; bias statistics such as the Draba (1977) statistic; and a copy of the test items and a separate answer key. The data should provide enough information for the committee to make an informed judgment. However, the committee does need to realize that field test data are not extremely accurate. The sample may not be representative, the motivation is less while taking a test that does not count, and individuals do not study in advance for a field test as they do

for an exam that counts, such as a licensure exam. However, the group of interns interviewed by the consultant group seemed quite motivated to perform well on the tests.

A useful set of data when attempting to validate the test based upon Georgia item tests would be the item data from the actual administration of the items in Georgia. These data would provide a set of comparison data for viewing item stability and possibly instructional differences.

If insufficient numbers of items exist at the end of the empirical validation process, the development of new items must be initiated; and the new items must be subjected to the same item validation process.

About 100 examinees are needed to make sound empirical validity judgments about field test items for a single test form. If at least 100 examinees are needed for development of a single test form, greater numbers of examinees (at least 200-300) are required for equating test forms than for determining item quality (Rentz and Rentz, 1978). Therefore, alternate forms may be desirable but unattainable.

Standard setting should be a multi-step approach to occur after the first real administration using the impact data from the test results and possibly from the field test.

Numerous methods exist for the establishment of standards on credentialing tests. Berk (1986) classified standard setting procedures as methods based (1) entirely on judgment (Judgmental), (2) primarily on judgment (Judgmental-Empirical), or (3) primarily on test data (Empirical-Judgmental). (Berk, 1986)

The method of setting standards is somewhat dependent upon the type of information that is available. The Judgmental method (Angoff, 1971) would be appropriate for a situation in which no examinee data are available for the standard setting process. The judges' content expertise and knowledge of the performance of the examinees are the only sources of information available. This method would be appropriate for tests such as Latin because there will be an insufficient amount of examinee data available to use in the standard setting process.

The Judgmental-Empirical data procedure is the one that is most often used in the development of standards for credentialing tests. Typically, the committee makes knowledge based judgments and are also provided with examinee data. A variation on this type of procedure involves asking teachers to judge the status of their students and having the students take the test (Garcia-Quintana & Huynh, 1980). The score for which there is the smallest number of inconsistent decisions (using the teacher

judgments and also examinee test data) is selected as the standard. This method is referred to as the contrasting groups procedure. The Empirical-Judgmental method relies primarily on performance data from one or more groups of examinees and the statistical analysis of that data. The Empirical-Judgment method requires too much examinee data to be used in setting standards on credentialing tests.

The Judgmental-Empirical method is the procedure most frequently used by institutions in the establishment of standards for credentialing.

A common procedure entails four stages (Saunders & Mappus, 1984). The first three steps require the committee members to make individual judgments from three differing points of view and to revise their personal opinions of the appropriate standard or passing score accordingly. In the final stage, the committee seeks to reach agreement on a recommended standard.

Before beginning the standard-setting process, the committee members are presented with a description of the context of the examination with particular emphasis on the examinees who take the test, a description of some of the technical characteristics of the test itself, and the requirements and procedures specified for the examination in the law or by the Board. This step is done to provide all committee members with a common background and to eliminate any misconceptions about the test. Next, the committee members are asked to conceptualize a group of hypothetical prospective individuals whose content knowledge place them at the borderline between those applicants who have achieved an adequate level of content knowledge for certification and those who do not. The judgments made by the committee members are based on their expectations of this hypothetical group.

The first stage of the standard setting procedure uses the (Angoff, 1985) method. For each item on the examination, committee members are asked to determine, individually, the percentage of the hypothetical borderline group that would be able to correctly answer each item. These data provide the information needed for step two of the procedure.

In the second standard setting step, staff members calculate a preliminary standard for each committee member by summing across the item judgments. This results in a preliminary total score standard for each committee member which is based on decisions made at the item level. These preliminary standards are presented to the committee members. Each member is then asked to consider the standard holistically, and, if deemed desirable, revise the preliminary standard to more accurately reflect what he or she feels the overall standard should be.

The third stage of the standard setting process requires consideration of the consequence of choosing a particular score for the standard. The committee is presented with an estimated distribution of the examinee scores. This distribution is derived from data obtained during an actual administration of the examination. Based on these data, each member is given the opportunity to again revise the standard.

Finally, the committee attempts to reach consensus on a single standard through discussion and voting. The first step in the consensus process is determining what level of agreement would represent committee consensus. The committee agrees on a consensus percentage which is usually a two-thirds or three-fourths majority.

During the discussion phase of the meeting, committee members express concerns about setting the standard either at the lower or upper end of the distribution. The consequences of a false negative or false positive decision are considered. A false negative decision is not certificating a person who may have adequate content knowledge; whereas, a false positive decision is certificating a person who does not have adequate knowledge.

In the final stage of the standard setting process, individual committee members are allowed to comment on the committee's recommendation in writing and to indicate if they could endorse the decision. The outlined procedure is not proposed as the <u>only</u> acceptable procedure. It is provided as an acceptable example.

After the final standard is recommended, the standard, along with the standard setting information, is presented to the Board for approval and/or adjustments in the standard prior to approval. The Board should specifically consider the relative costs of false positives and false negatives. This may result in their adjusting the standard either up or down.

<u>Given the technical nature of the test development process, the further development of the DCPS tests be assigned to a unit with appropriate research and measurement expertise while maintaining continued collaboration of the Intern/Mentor Program. This includes supervision of the NES contract and coordination of any external monitoring.</u>

DCPS can contract with NES to develop a test for licensure purposes, but the responsibility of assuring test quality can not be abdicated. Test quality can only be assured by having technically qualified individuals overseeing the test development process. We are certainly not suggesting by this that NES is not a professional organization and that they would not do a quality job without technical supervision. Nevertheless, DCPS cannot

allow NES to be their own technical supervisor. Experience in other states suggests that technical supervision is preferred to non-technical supervision.

<u>There should be established an ongoing technical advisory committee.</u>

The rationale for this recommendation has basically been stated under recommendation #2 above. There are periodic times in any test development process where judgments need to be made regarding the adequacy of completed steps or regarding the specific design of subsequent procedures. No staff contains the total amount of wisdom regarding test development for teacher licensure test, but considerable wisdom does exist across the country. For a high stakes test such as this one, it is simply prudent to obtain outside advice regarding technical issues. It is the experience of the writers of this report that such outside advice is very valuable. Indeed, the technical advisory committees in various states do not limit their advice to technical issues but give useful advice on policy issues also.

We suggest that a three or four member technical advisory committee be established. This set of individuals should be ones who are well recognized by their colleagues as being experienced in and having expertise about licensure testing. Such individuals could come from either State Departments of Education or University faculties. We suggest it may be wise to have members from both types of settings. The advisory committee should meet periodically, but there probably should not be meetings set on a regular basis. The meetings should only take place when specific decisions are to be made for which the local staff requires some external advice.

The recommendations discussed above are intended to result in the documented evidence of validity necessary to support the use of the DCPS subject matter tests for teacher certification purposes. In our view, the documentation that is available currently is not sufficient for this purpose. The recommendations are offered as an example of the type of additional actions that are necessary, rather than as a rigid, step by step prescription. If actions of the nature described above are taken within a policy environment such as we have recommended, we expect DCPS to soon be in a position to utilize their subject matter tests as an additional criterion of teacher competency.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME) Joint Committee. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational Measurement (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. Review of Educational Research, 56, #1, 137-172.

Cronbach, L.J. (1980). Validity on parole: How can we go straight? In W.B. Schrader (Ed.) Measuring Achievement: progress over a decade. New directions for testing and measurement (pp. 99-108). #5. San Francisco: Josey Bass.

Draba, R. E. (1977) The identification and interpretation of item bias. Memorandum No. 26, Statistical Laboratory, Department of Education, University of Chicago.

Garcia-Quintana, R. A., & Huynh, H. (1980, March). Setting passing scores in the readiness test for beginning first graders in the state of South Carolina: some practical considerations. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Lawrence Johnson & Associates, Inc. (1985, July). Content inventory analysis: Final report. Washington, DC: author.

NES (1987), October). Interim content knowledge assessment program final report. Amherst, Mass.: author

Rentz, R.R. & Rentz, C.C. (1978). Does the Rasch model really work? A discussion for practitioners. Princeton: Eric Clearinghouse on Tests, Measurement and Evaluation.

Saunders, J.C. & Mappus, L.L. (1984, April). Accuracy and consistency of expert judges in setting passing scores on criterion-referenced tests: The South Carolina experience paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Traub, R.E. (1986). Review of test reliability assessment procedures and results. In R.M> Jaeger and J.C. Busch (Eds.). An evaluation of the Georgia Teacher Certification testing program. (Chapter 5). Greensboro, NC: Center for

Educational Research and Evaluation, University of North Carolina.

APPENDIX A

OVERVIEW OF LEGAL ISSUES IN TEACHER TESTING

OVERVIEW OF LEGAL ISSUES IN TEACHER TESTING

I. Introduction

As part of the much-publicized effort to improve education in the United States, many states have adopted requirements that teachers -- usually entry-level teacher candidates -- demonstrate some level of competence, measured by a written examination, before being allowed to teach in the public schools. Given the litigious nature of modern American society, the arguably-harsh result of disqualifying highly-educated persons on the basis of a single pencil and paper test, and the unfortunate reality that members of racial minority groups frequently perform relatively poorly on written tests, it is reasonable to assume that some of these teacher testing programs will be challenged in court. Teachers or teacher candidates who fail to pass the examinations and are denied jobs or certification as a result can be expected to sue, perhaps in major class action litigation. No school district should implement a teacher testing program without first having come to grips with the real possibility that its testing program will eventually be the subject of a legal challenge, and without having planned in advance for that possibility. This document will discuss, in very general terms, some of the most obvious legal issues which might arise in such litigation, and

will make some suggestions about advance preparation for dealing with such litigation.[1]

II. The Constitutional and Statutory Foundation for Teacher Testing Litigation

Although creative lawyers representing disappointed teacher candidates may come up with some novel legal theories upon which to base a challenge to a teacher testing program, it is predictable that one or more well-recognized constitutional and statutory theories will be advanced. Two of the theories discussed below are based on race discrimination, while the others are Constitution-based theories not dependent upon race.

A. Title VII of the Civil Rights Act of 1964. 42 U.S.C. §2000e, et seq., prohibits employment discrimination on the basis of race, sex, or ethnic group.[2] Employment tests clearly

---

1. These comments give a broad overview of the subject, and are not intended to be legal advice to DCPS or the Superintendent. Particularly because of the "city - state" nature of the Federal District, some of the principles discussed herein may not be fully applicable to DCPS. Legal counsel for DCPS should be consulted in connection with all legal issues pertaining to the Teacher Testing and Assessment Program.

2. It is assumed that DCPS is an employer, at least for some purposes, within the meaning of Title VII. DCPS legal counsel should be consulted on this issue.

are "an employment practice" covered by this statute. In the most typical litigation situation, tests are challenged because they have an adverse or disparate impact on minority applicants. The legislative history of Title VII evidences an awareness of the increasing use of employment tests and of the fact that, in some instances, unreliable and biased tests were being used to deprive minorities of fair employment opportunities.

Proof of a Title VII violation involving an employment test has three parts. First, the plaintiff (normally a plaintiff group or class) has the burden of demonstrating that the test has "adverse impact" on a protected group. If the plaintiff carries this burden, the burden then shifts to the defendant to show that its test is "job related." Finally, if the test is shown to be job-related, the plaintiff has an opportunity to prove that other tests or selection devices without a similar undesirable racial effect would serve the defendant's business purpose just as well as the challenged test. See, e.g., Griggs v. Duke Power Co., 401 U.S. 424 (1971); Albermarle Paper Co. v. Moody, 422 U.S. 405 (1976). As a practical matter, most cases have been decided based on the first two steps.

The requirement that the plaintiff first show that the test in question has an adverse racial impact has been phrased in varying language. In Griggs, the Court said a test must "operate to disqualify Negroes at a substantially higher rate than white applicants." 401 U.S. at 425. The Albermarle Court spoke of tests which select applicants for hire or promotion "in a racial

pattern significantly different from the pool of applicants." 422 U.S. at 425. As a "rule of thumb", it has often been said that adverse impact exists if the passing rate for blacks is less than 80% of the passing rate for whites. The federal courts have used this "4/5ths rule" and have tended also to examine whether or not there is a statistically significant difference between the respective pass rates. There is a large body of case law on this subject.

If the plaintiff demonstrates adverse impact, the burden is on the defendant to justify its test on grounds of business necessity. Normally, this burden is carried by validation which demonstrates the relation between the selection procedure (test) and performance on the job. The Uniform Guidelines on Employee Selection Procedures, 29 CFR section 1607, contain comprehensive standards for determining the sufficiency of validation studies of employment tests. The Uniform Guidelines are frequently given great weight by the federal courts. See, e.g., Albermarle, 422 U.S. at 341. In addition, the Standards (AERA-APA-NCME, 1985) contain a chapter specifically dealing with validation of employment tests.

As noted, Title VII and the Uniform Guidelines apply to employment tests. They do not apply to all tests, and there is a very strong argument, discussed in more detail below, that they do not apply to licensing tests. A discussion of the nature of licensing tests is important to understand the legal distinction between them and employment tests.

Licensure is the process by which an agency of government

grants permission to an individual to engage in a given occupation or profession upon finding that the applicant has attained the minimal degree of competency required to ensure that the public health, safety, and welfare will be reasonably well protected. That is, the purpose of licensure is to protect the public. To become licensed, applicants frequently are required to meet specified requirements as to education, training, or experience. Further, licensing agencies generally rely on tests of competence to determine those applicants who have met a standard and who deserve to be granted a credential.

In Goldfarb v. Virginia State Bar, 421 U.S. 773 (1975), the Supreme Court said as follows:

> (T)he States have a compelling interest in the practice of professions within their boundaries, and that as part of their power to protect the public health, safety, and other valid interests they have broad power to establish standards for licensing practitioners and regulating the practice of professions.

421 U.S. at 792.

In the same way, DCPS has a compelling interest in determining whether persons seeking to teach have the minimum content knowledge necessary to teach in the classrooms of the District of Columbia. It is entirely legitimate for DCPS to use an examination for this purpose. As stated in Tyler v. Vickery, 517 F.2d 1089 (5th Cir. 1975), cert. denied, 426 U.S. 940 (1976):

> If a state has the right to insist on a minimum standard of ... competence as a condition of licensure, it would seem to follow a fortiori that it may require a demonstration of such competence in an examination...

517 F.2d at 1101.

The tests and other procedures used in assessing competence for licensing are in most respects similar to those used for evaluating educational achievement and for making employment decisions. However, there are certain aspects of licensing testing that differ substantially from those found in educational testing and personnel selection testing.

Since licensing tests are used to determine an applicant's fitness (at a minimum competence level) for practice, it is generally recognized that it is essential that such a test be job-related. The procedures used for identifying critical or important job-related knowledge, skills and abilities are largely the same as those used for developing employment tests. There are, nevertheless, sharp differences in purposes between employment tests and licensing tests. Employment tests are designed to identify individuals in an applicant group who are most likely to be successful on the job or in training. Their purpose is to predict job <u>success</u>. The purpose of licensing, on the other hand, is to protect the public health, safety, and welfare. For this reason, tests used for licensing must be able to help identify those who possess the knowledge, skills and abilities deemed necessary to safeguard the public. Licensing tests are <u>not</u> intended to predict job success. That is, the purpose of licensing testing is to assess certain knowledge or skills which are critical to the protection of the public, but not to test the wide range of other knowledges, skills, and

abilities which one might need to have to actually be successful in the occupation or profession in question. In other words, licensing tests are designed to test "necessary" -- but not "sufficient" -- knowledge and skills.

From this basic difference in purpose, it follows that tests designed for licensing will differ from those designed for meeting the selection or promotion needs of an employer. Certain types of knowledge and skills that may be very important to job success might not be appropriate for inclusion on a license exam. For example, a pleasing personality and sales ability are usually considered important attributes for success as a real estate sales person. Anyone developing a test to select personnel for this occupation would almost certainly strive to assess these qualities in some way. However, since the licensing agency is not concerned with predicting job success, it would probably ignore such attributes (even though they may be job-related) in favor of knowledge and skill directly related to the purpose of licensing, which is to protect the public.

Another way of pointing out the difference between employment tests and licensing tests is this: Generally, employment tests are designed to identify and select the most qualified applicants for the job in question. Consequently, these tests frequently include questions or problems that will challenge even the most highly qualified applicants. On the other hand, licensing examinations are designed to assess minimum competency, and are <u>not</u> designed to differentiate between or among people beyond the minimum competence level.

Because licensing tests are not designed to select the "best" of the applicants, but rather are designed to assess whether each applicant is above or below the minimum standard, licensing tests should most appropriately use an absolute standard. That is, whether a particular individual passes or fails the examination should not be influenced by how that applicant performs relative to the rest of the applicants taking the test. Rather, the applicant should be judged by how he does against the pre-determined standard of minimum competence. By using an absolute standard, all the applicants theoretically could pass or all the applicants could fail. Thus the performance of any one applicant could in no way influence whether another applicant was licensed or not. This method of testing and standard setting is generally referred to as a criterion-referenced testing approach.

The distinction between employment testing and licensing testing is of sufficient importance that the 1985 Standards discuss employment testing and licensure testing in separate chapters. These Standards are generally regarded as one of the principal authorities in the profession regarding the development and use of tests.

In the chapter on licensing testing, the Standards say as follows:

> The primary purpose of licensure or certification is to protect the public. Licensing requirements are imposed to ensure that those licensed possess knowledge and skills in sufficient degree to perform important occupational activities safely and effectively.

* * *

> Issues of validity that are discussed in other sections of the standards are also relevant to testing for licensure and certification. Although many of the issues of central importance in the present context are discussed in the chapter on employment testing, some important distinctions must be made. For licensure or certification the focus of test standards is on levels of knowledge and skills necessary to assure the public that a person is competent to practice, whereas an employer may use tests in order to maximize productivity.

* * *

> Although the job analysis techniques are comparable to those used in employment testing, the emphasis for licensure and certification is limited appropriately to knowledge and skills necessary to protect the public.

* * *

> Skills that may be important to success but are not directly related to the purpose of licensure (i.e., protecting the public) should not be included in a licensing exam.

Several federal courts which have considered the question have concluded that an agency engaged in licensure decisions is not making employment decisions, and consequently that Title VII and the Uniform Guidelines do not apply to licensing tests. See, e.g., Tyler v. Vickery, 517 F.2d 1089, 1096 (5th Cir. 1975); Haddock v. Board of Dental Examiners of California, 777 F.2d 462 (9th Cir. 1985); Woodward v. Virginia Board of Bar Examiners, 598 F.2d 1345, 1346 (4th Cir. 1979); George v. New Jersey Board of Veterinary Medical Examiners, 794 F.2d 113 (3rd. Cir. 1986).

If Title VII and the Uniform Guidelines do not apply, then the more rigid requirements for establishing validity for employment tests do not apply. Instead, validity can be

established and demonstrated through the more flexible and more realistic content-validity approach discussed in the chapter on licensing testing in the 1985 Standards and in the professional literature on licensing testing. Therefore, it is important that DCPS be clear and consistent in the design and use of its test about its purpose as either a licensure test or an employment test.

The DCPS consultants have recommended testing be used as a licensing device not as an employment device. One important caveat is in order here. Unually, the agency making licensure or certification decisions does not play any role in making subsequent employment decisions with respect to individuals holding the credential. However, DCPS acts both as a licensing agency for teachers, and as an employer of teachers. Generally speaking, one cannot be employed over the long term by DCPS as a teacher without having first been licensed by DCPS. This unusual situation may make it more difficult for DCPS to contend that its test (as has been recommended) is being used as a part of its licensing activities, and not as an employee selection device. This potential problem reinforces the need for DCPS to take great care in the development and use of its test as part of its licensing function, to collect validity evidence in a manner consistent with a licensing test, and to be clear and unequivocal about the purpose of its test.

B. Title VI of the Civil Rights Act of 1964. Title VI, 42 U.S.C. Section 2000d, et seq. proscribes discrimination on the basis of race in "any program or activity receiving federal

financial assistance." Title VI may or may not apply to the DCPS testing and assessment program, depending upon whether that program itself receives federal financial assistance. See Grove City College v. Bell, 465 U.S. 555 (1984). DCPS legal counsel should be consulted on this point.

If Title VI does apply, it could form the basis of a legal challenge to the DCPS teacher testing program. The legal analysis would be essentially the same as for a Title VII claim, but applicability of the statute would not depend upon whether DCPS is an employer or engaged in employment testing. That is, Title VI could apply to a licensing testing program. In Guardian's Assn. v. Civil Service Commission of New York City, 463 U.S. 582 (1983), the Supreme Court held that Title VI not only proscribes intentional racial discrimination, but also reaches actions which have a discriminatory effect even though not intentionally based on racial considerations. The Court went on to explain that the order of proof model developed for employment discrimination cases is applicable to Title VI disparate impact claims as well. In other words, once the plaintiff establishes adverse impact, the recipient of federal funding must prove a "business necessity" for the practice having a discriminatory impact. See Georgia State Conference of Branches of NAACP v. State of Georgia, 775 F.2d 1403 (11th Cir. 1985) (applying Title VII disparate impact analysis to Title VI claim over alleged discriminatory student assignment practices).

It seems clear that the defendant's burden of establishing "business necessity" can be carried by introduction of

appropriate evidence of the test's validity. If the test is a licensing test, evidence that it has been properly validated for that purpose should carry the defendant's burden. That is, the defendant, through this validity evidence, will be able to show that there is a demonstrable relationship between the test and the defendant's duty to protect the public through insuring that classroom teachers have knowledge of the subject matter which they will teach.

C. The Constitution. Individuals challenging a test required for licensure may make a Constitution-based challenge -- in the case of states under the Fourteenth Amendment, and in the case of DCPS under the Fifth Amendment. For purposes of this discussion, it will be assumed that the legal analysis would, for all practical purposes, be the same under both amendments. See, e.g., Bolling v. Sharpe, 347 U.S. 497 (1954). Legal counsel for DCPS will be in the best position to assess the accuracy of this assumption.

The Supreme Court held in Washington v. Davis, 426 U.S. 229 (1976), that state laws or regulations cannot be invalidated under the Fourteenth Amendment solely because of their differential impact on persons of different racial groups. Rather, it is necesary that the state action be proven to have a discriminatory purpose. In the absence of evidence of such purpose, the Constitution requires only that the state's use of the law or regulation (or test requirement) be rationally related to the accomplishment of a legitimate governmental

objective.[3/] See Tyler v. Vickery, 517 F.2d 1089 (5th Cir. 1975), cert. denied, 426 U.S. 940 (1976).

There can be no doubt that DCPS would be pursuing a legitimate governmental objective in trying to assure that new teachers have at least a certain level of subject matter knowledge before being licensed to teach. If a licensing test requirement is the device chosen to attain this objective, DCPS must simply be able to demonstrate that there is a logical and rational nexus between the knowledge measured by the test and knowledge required on the job. This demonstration can be made through presentation of the content validity evidence developed through the test construction process.

Another constitutional issue could arise. In Debra P. v. Turlington, 474 F.Supp. 244 (M.D. Fla. 1979), aff'd in part and rev'd in part, 664 F.2d 397 (5th Cir. 1981), on remand, 564 F.Supp. 177 (M.D. Fla. 1983), aff'd, 730 f.2d 1405 (11th Cir. 1984), the federal courts discussed various constitutional issues arising in a student testing context. One of these -- curricular or instructional validity -- is discussed briefly in the next section. Another of the issues warrants mention here.

The courts concluded in Debra P. that it would violate students' due process rights to implement a graduation test requirement without reasonable notice and a phase-in period. This notion of "fundamental fairness" could be applied as well in

---

3. Although some teacher testing programs have been challenged on grounds including intentional racial discrimination, there is no reason to expect such a challenge of a DCPS test.

the context of a teacher certification or licensing test. A court might well insist that reasonable notice be provided to teacher candidates and that the candidates be given reasonable warning of changed standards and adequate time to prepare for any new tests. However, the countervailing need to protect the public from unknowledgeable teachers would probably allow DCPS to minimize the length of any phase-in period.

D. Curricular and Instructional Validity. A plaintiff might argue, based on a loose interpretation of Debra P., supra, that the DCPS test cannot be used unless DCPS can demonstrate that the test has "curricular" or "instructional" validity. In Debra P., the court held that in order to insure "fundamental fairness", Florida could not include -- on a test required for graduation -- material which had not been taught to all students required to take the test.

In a teacher licensing context, a disappointed candidate might claim that his or her constitutional rights have been violated because the licensing test covers materials and subject matter not taught in the teacher preparation program which trained the teacher candidate. While it might be advisable to consider the content of teacher preparation programs when defining the domain to be tested in a teacher licensing program, it seems clear that the vague, largely non-psychometric notions of curricular and instructional validity have no legal significance in a licensing context.

Licensure tests are designed to protect the public and, for

this reason, the appropriate judgment of validity should be based on whether or not the tests cover the knowledges and skills that those licensed should possess. For the purpose of the licensure decision, it is irrelevant and inappropriate to consider curricular validity, i.e., whether the test matches the teacher training curriculum, in judging the quality of the test.

While the notion of instructional or curricular validity may be important if one wishes to make inferences about instructional effectiveness, and is important for minimal competency tests for high school graduation, it is wholly irrelevant to the quality of a licensure examination, the purpose of which is to protect the public from incompetent practitioners.

The 1985 Standards implicitly recognize the legitimacy of the distinction between student testing and licensure testing. Although they do not use the term "curricular validity," they do address the notion in Chapter 8, "Educational Testing and Psychological Testing in the Schools." Chapter 11, "Professional and Occupational Licensure and Certification," however, makes no mention of such a standard.

For additional information, see Davidson v. State of Georgia, 622 F.2d 895 (5th Cir. 1980) (there is no federal statutory or constitutional right to be taught how to pass the bar examination).

III. Conclusion.

There is legal as well as psychometic significance for the recommendation that the DCPS teacher tests be developed and used for licensing purposes rather than employment purposes. There are substantial differences between licensure tests and employment tests in terms of establishment and demonstration of validity. Proof of validity of a licensure test would probably not be subject t o the rigors of the somewhat-anachronistic Uniform Guidelines as would an employment test. Rather, the 1985 Standards would supply guidance for a court's consideration of the validity questions.

Regardless of the nature of the legal challenge to the teacher tests, DCPS should assume that it will be called upon in court to prove the validity of the tests. Proof of validity will require accurate and thorough documentation of all activities, and the results of all activities, in the test development process. It will also require careful and consistent application of the tests to their intended purpose.

APPENDIX B

REPORT ON TEST DEVELOPMENT AND INTERN PERFORMANCE

Appendix B

Report on Test Development and Intern Performance

on the

Intern Content Knowledge Assessment Instruments

The Intern Content Knowledge Assessment Instrument consists of 12 subject-area tests and a writing test. The subject area tests have been developed for the following areas:

1. Early Childhood Education
2. Elementary Education
3. English
4. Social Studies
5. Mathematics
6. Science
7. English as a Second Language
8. Special Education
9. French
10. Spanish
11. German
12. Latin

The objectives and test items for each of those subject areas were rated by experienced teachers from the District of Columbia Public Schools (DCPS) at a Validity Conference organized by the test developer, National Evaluation Systems (NES). The numbers of teachers providing ratings for each subject area are shown in Table 1 below.

TABLE 1

NUMBERS OF TEACHERS PROVIDING RATINGS OF OBJECTIVES AND ITEMS BY SUBJECT AREA

Early Childhood Education	11
Elementary Education	14
English	12
Social Studies	10
Mathematics	10
Science	10
English as a Second Language	5
Special Education	11
French	5
Spanish	8
German	1
Latin	1

The process used by the Contractor to determine the content validity of the objectives and items are those which are generally accepted within the measurement community. Whether that process has been applied with sufficient rigor will require further review. In rating objectives, the participants were asked if they were familiar with the content reflected by a given objective. If they were, raters were then asked to indicate whether or not the objective was valid according to the following criteria:

1. Level of Knowledge: Is the Objective at an appropriate level of knowledge for an entry-level educator?

2. Accuracy: Is the content of this objective accurate?

3. Free of Bias: Is the objective free of content and language that offends or disadvantages examinees on the basis of any personal characteristics?

4. Represents Knowledge Used to Perform the Job of an Entry-Level Educator in DCPS: Is the material contained in this objective representative of the knowledge that an educator would use to perform the job of an entry-level educator in DC Public Schools? (from the Draft Final Report, p.10)

If an objective was rated as invalid, raters were asked to specify which of the above four criteria was not met.

A similar process was used in rating items. After determining if the rater was familiar with the content of the item, the rater then evaluated the item on the following criteria:

1. Objective Match: Does the item measure an important aspect of the objective?

2. Accuracy: Is the content of this item accurate and is there one correct or best answer designated?

3. Free of Bias: Is the item free of content and language that offends or disadvantages examinees on the basis of any personal characteristics?

4. Represents Knowledge Used to Perform the Job of an Entry-Level Educator in DCPS: Is the material contained in this item representative of the knowledge that an educator would use to perform the job of an entry-level educator in DC Public Schools (from the Draft Final Report , p.11).

Participants were also asked to provide standard setting information for every item considered valid. They were asked to answer the following question about a hypothetical group of new teachers:

> "What percentage of entry-level educators who have the minimum amount of content knowledge necessary to teach acceptably in DCPS in their certification field would answer the item correctly? (Draft Final Report. p.11)"

Based on the ratings provided by DCPS teachers at the Validity Conference, 4 of 650 objectives were discarded and 1 item was considered not valid. That item and the items related to the four objectives were removed from the item bank. The Project Director from NES has stated that the content validation process has not been completed for the language subject area tests. Additional teachers will be asked to rate the validity of the objectives and items in those tests.

The subject area tests have been administered to DCPS interns who volunteered for testing on four separate occasions. The first three adminstrations were considered field tests of the instruments, although the third administration (May, 1987) used a version of the instrument identical to the actual test which was administered in September, 1987. The numbers of interns who have taken the tests are shown by subject area in Table 2.

TABLE 2

NUMBER OF INTERNS WHO HAVE VOLUNTARILY TAKEN THE INTERN CONTENT KNOWLEDGE ASSESSMENT INSTRUMENTS

		Number of Interns Tested				
		5/86	9/86	5/87	9/87	Total*
01	Early Childhood	19	36	32	14	201
02	Elementary Education	11	22	24	47	104
03	English	0	2	3	12	17
04	Social Studies	0	2	1	2	5
05	Mathematics	4	8	6	16	34
06	Science	10	12	6	2	30
07	English as a Second Lang.	0	7	1	9	17
08	Special Education	16	23	18	14	71
09	French	1	2	2	1	6
10	German	0	0	0	0	0
11	Latin	0	0	0	0	0
12	Spanish	1	4	4	11	20
		62	118	97	128	405

* Only the final test administration of September, 1987 used the actual version of the tests, although the May 87 field test version is considered identical to the actual test.

It should be noted that the numbers of interns who have taken these tests is too small to make definitive statements concerning the performance of DCPS interns on the tests; however, given that qualifer, we will examine the 1987 administrations of the tests for indications of intern performance.

During Spring of 1987, there were 196 interns in the Intern-Mentor Program. Ninety-Seven (49%) of those interns voluntarily took the subject area tests and 93 took the writing test. The numbers of interns in the program and those taking the tests, by subject area, are shown in Table 3.

TABLE 3

NUMBER OF INTERNS IN THE INTERN/MENTOR PROGRAM
AND NUMBERS TAKING THE SUBJECT AREA AND
WRITING TESTS IN MAY,1987

SUBJECT AREA	N INTERNS	N SA TEST	N WRITING TEST
Early Childhood	36	32	32
Elementary Ed	44	24	22
English	6	3	3
Social Studies	6	1	1
Mathematics	13	6	5
Science	23	6	6
ESL	11	1	0
Special Education	44	18	18
French	4	2	2
Spanish	9	4	4
German	0	---	---
Latin	1	0	0

Each of the subject area tests consists of 120 items, 100 of which are scoreable. The score that is reported to the interns by the Contractor is the percentage of items correct for subtests within the test, and for the entire test. The number of subtests ranges from a low of two for French

to a high of six for Early Childhood, Elementary Education, and English. The other tests have five subtests.

For the May, 1987 administration of the writing test, 8 separate areas were scored reflecting the scoring criteria and procedures that were approved by the DC Writing Committee in June, 1987. The areas were:

1. Organization
2. Audience and Purpose
3. Word Choice
4. Sentence Structure
5. Main Ideas and Details
6. Transitions
7. Grammar and Usage
8. Mechanics

Scores of 0, 1 or 2, were assigned to each area for the intern's writing exercise indicating that the intern's writing failed to meet minimum standards (0), met minimum standards (1) or exceeded minimum standards (2). Scores for interns, then, could range from a low of 0 to a high of 16. The Writing Field Test Report indicated that the scoring protocol will change in future administrations. Holistic scoring will be used and scores ranging from a low of 1 (little control of essential writing skills) to 4 (very strong control of essential writing skills) will be assigned to intern writing samples by two raters. Scores will then range from 2 to 8.

An examination of Table 3 shows that only three of the subject area tests and the writing test have large enough n's for any further discussion. The means and ranges of scores for each of those tests is shown in Table 4.

TABLE 4

MEANS AND RANGES FOR THREE SUBJECT AREA TESTS AND THE WRITING TEST. MAY, 1987

AREA	N	MEAN	RANGE
Early Childhood	32	73.6	51-88
Elementary Education	24	69.8	34-90
Special Education	18	71.4	51-93
Writing	93	9.3	1-16

NES is presently in the process of determining passing scores based on the ratings from the Validity Conference and intern performance. However, for this exercise, four potential passing scores of 65, 70, 75, and 80 were chosen to determine the intern passing rates for the subject area tests. The percentage of interns who would have passed the tests for each of those passing scores is shown in Table 5.

TABLE 5

PERCENTAGES OF INTERNS PASSING SUBJECT AREA TESTS FOR DIFFERENT PASSING SCORES MAY, 1986

	65%	70%	75%	80%
Early Childhood	90	68	50	37
Elementary Education	66	58	45	29
Special Education	72	50	38	27

There are three additional subject area tests with multiple examinees. Two of those tests--Mathematics and Spanish--had mean scores similar to those reported above (77.3 and 64.7, respectively). However, the Science

test had a mean of 49.0 with a range of 28 to 71 percent correct. It should be noted that 23 interns took an earlier version of the subject area test in May, 1986. For 21 of those interns, the score received in the 1987 administration was higher than the 1986 score. Those results are presented below.

TABLE 6

INTERN SUBJECT AREA TEST SCORES
MAY, 1986 AND MAY, 1987

SUBJECT	5/86 SCORE	5/87 SCORE	DIFFERENCE
01	75	80	+ 5
02	75	87	+12
03	61	78	+17
04	70	84	+14
05	67	67	0
06	63	74	+11
07	48	65	+17
08	60	79	+19
09	59	69	+10
10	59	69	+10
11	65	71	+ 6
12	66	93	+27
13	58	63	+ 5
14	38	46	+ 8
15	32	28	- 4
16	37	51	+14
17	59	69	+10
18	58	71	+13
19	51	64	+13
20	78	93	+15
21	57	63	+ 6
22	65	76	+11
23	55	69	+14

The differences between the 1986 and 1987 administration of the tests ranged from a low of -4 to a high of +27 with an average difference of 11 percentage points. Tests were re-administered to interns in six of ten subject areas--elementary education, early childhood education,

Spanish, mathematics, special education, and science. If 65% were designated as the required passing score, 8 interns (35%) from the May 86 administration received the necessary score. However, one year later, 17 interns (74%) received that minimum score. An examination of the remaining six interns shows that two of them were within two percentage points of that score and the remaining four were all science interns (The science test covers the areas of general science, earth science, chemistry, biology, and physics, rather than just the area in which certification is sought. Intern scores have been considerably lower on this test than on the other subject area tests).

An additional exercise was attempted--to look at a passing score on the subject area test in combination with a score on the writing test. We arbitrarily chose the score of 70% on the subject area test in combination with a score of 8 on the writing test. Using those two criteria, the passing rates for the three subject area tests above are shown below:

Early Childhood	53%
Elementary Education	45%
Special Education	27%

Given the current literature concerning the differential passing rates of Black and White teachers on teacher assessment instruments, we attempted to examine the passing rates of White and Black DCPS interns. Only four interns taking the subject area tests were identified as non-Black. The White interns received the highest score for two of the subject area tests, and the second highest for the third test. The remaining White intern did not score highest on the Subject Area test, but did score above the mean of those tested.

In late September, 1987, 128 of the 199 interns in the DC Intern-Mentor Program, voluntarily took the subject area tests and the writing test. The numbers of interns who participated, by subject area, are presented in the Table 6 below:

TABLE 6

NUMBER OF INTERNS IN THE INTERN/MENTOR PROGRAM AND NUMBERS TAKING THE SUBJECT AREA TESTS IN SEPTEMBER, 1987

AREA	#INTERNS	#TESTED	PERCENT
Early Childhood	21	14	67
Elementary Education	65	47	72
Special Education	23	14	60
Mathematics	23	16	70
Science	19	2	11
Spanish	14	11	79
English	15	12	80
French	1	1	100
Social Studies	6	2	33
English as a Second Lang	12	9	75
German	0	--	--
Latin	0	--	--
TOTAL	199	128	64

The results of the May and September test administrations are presented below in order to compare intern performance on the two administrations and to examine the percentage of interns who would have passed the tests for selected passing scores. All results are reported when

n is 9 or larger for the September administration. *N* for the each administration is indicated in brackets beside the mean score. Again, the number of examinees in each of the subject areas is too small to make definitive judgements.

TABLE 7

COMPARISON OF SPRING, 1987 AND FALL, 1987
TEST RESULTS FOR SELECTED SUBJECTS

AREA	MEAN 5/87 (range)	MEAN 9/87 (range)	PASSING RATE BY SCORE: 65%	70%	75%	80%
ECE	74 [32] (51-88)	68 [14] (54-82)	71%	36%	7%	7%
El Ed	69 [24] (34-90)	69 [47] (27-95)	64	45	38	28
Sp Ed	71 [18] (51-93)	64 [14] (44-82)	57	36	21	7
Math	77 [6] (43-93)	56 [16] (4-92)	44	38	25	13
Spanish	64 [4] (31-85)	63 [11] (20-95)	55	45	45	45
English	73 [3] (64-93)	56 [12] (18-90)	25	25	25	16
ESL	[1]	79 [9] (54-96)	89	78	56	44

Each of the interns taking a subject area test submitted a writing sample giving an *n* of 128 for the writing test. The range of possible scores on this test was 2 through 8, which is the range reported for these interns. The distribution of writing scores is shown in Table 8 below:

TABLE 8

DISTRIBUTION OF WRITING SCORES FOR THE SEPTEMBER,1987 INTERN WRITING SAMPLES

SCORE	FREQUENCY	PERCENT
Not Scorable	4	3
2	10	8
3	9	7
4	31	24
5	24	18
6	28	21
7	14	11
8	8	6

While no 'passing score' has been calculated, a passing score of 5.0, the mean for the 124 writing samples that were scored, was implied. All papers scored as 5 or below were diagnostically scored to determine areas of weakness. 74 of 124 papers scored at 5 or below (60%). The numbers of interns below standard, by area, are shown below.

1.	Present an introduction, body, and conclusion related to the topic	31
2.	Present main ideas and details in a logical sequence	32
3.	Use appropriate transitional words, phrases and sentences	16

4.	Address the specified audience and purpose	21
5.	Choose effective and precise words	26
6.	Compose clear and complete sentences	33
7.	Employ standard written English grammar and usage	30
8.	Use mechanics of standard written English	38

The numbers above indicate that interns were generally below standard in more than one area. Table 9 shows the number of areas in which interns scored below standard.

TABLE 9

DISTRIBUTION OF NUMBERS OF AREAS BELOW STANDARD IN THE WRITING SAMPLE SEPTEMBER, 1987

NUMBER OF AREAS BELOW STANDARD	Frequency	Percent
1	1	1
2	21	28
3	28	38
4	17	23
5	5	7
6	2	3
TOTAL	74	100

Finally, an analysis of the test results was conducted to determine whether there were differences in test scores for Black and White interns. Again, it must be stressed, that with one exception, the numbers of interns

in any specific subject area are too small to make definitive statements about the observed differences in test scores. However, the results support the trend reported for the previous DCPS administration of these tests and support the national trends.

There are six subject areas in which Black and White interns were tested. In the areas of Special Education and English, the scores of the White interns are higher than the scores of all Black interns who were tested. For the subject areas of Mathematics and Spanish, the scores of the White interns are not the highest scores, but they are at least 20 percentage points higher than the mean scores for all interns tested in those areas. In the area of English as a Second Language, the majority of interns tested were White. For this test, the score of the Black intern is below the mean for all interns who took this test.

In Elementary Education, there are an almost equal number of Black and White interns, 27 and 20 respectively. We examined this area to determine potential 'passing rates' for Black and White interns, given two different standards. The overall mean for this area is 69%. The mean for Black interns is 59.8%, while the mean for White interns is 81.5%. The mean score of 69% was selected as one possible "passing score" and 64% (2 standard errors below the overall mean) was selected as a second one. The results are shown in Table 10.

TABLE 10

PASSING RATES FOR BLACK AND WHITE INTERNS IN ELEMENTARY EDUCATION FOR TWO SELECTED PASSING SCORES

SCORE	TOTAL	WHITE INTERNS	BLACK INTERNS
69	57%	95% (19/20)	30% (8/27)
64	68%	100%	44% (12/27)

Epilogue

The information presented above represents the data which have been collected to date concerning the technical properties of the Intern Content Knowledge Assessment Instruments and the performance of the DCPS interns on those tests. This information was considered by the DCPS Consultants and the Superintendent's Special Task Force in shaping the recommendations concerning future test development and policies for the use of the tests. Technical monitoring and analysis of test results will continue as part of the test development process.

APPENDIX C

CONSULTANT BRIEFING PAPER

PROCEDURES AND OBJECTIVES FOR CONVENING NATIONAL EXPERTS IN RESEARCH AND PRACTICE TO:

ASSESS THE DEVELOPMENT OF AND MAKE RECOMMENDATIONS FOR THE IMPLEMENTATION OF A BEGINNING TEACHER TESTING AND ASSESSMENT PROGRAM IN THE DISTRICT OF COLUMBIA PUBLIC SCHOOLS

Michael Kane
Barbara Williams, Co Chairmen

October 30, 1987

PROCEDURES AND OBJECTIVES FOR CONVENING NATIONAL EXPERTS IN RESEARCH AND PRACTICE TO: ASSESS THE DEVELOPMENT OF AND MAKE RECOMMENDATIONS FOR THE IMPLEMENTATION OF A BEGINNING TEACHER TESTING AND ASSESSMENT PROGRAM IN THE DISTRICT OF COLUMBIA PUBLIC SCHOOLS.

INTRODUCTION

During mid November a small, diverse group of individuals will spend three days interacting with each other and with officials and staff of the District of Columbia Public Schools (DCPS). These "external" professionals have in common in-depth experience with matters related to the development and implementation of testing programs for educational personnel, particularly beginning teachers. They are being asked to join DCPS officials in what is known as a "convening process" in order to assist DCPS in determining how to best implement its beginning teacher testing and assessment program. This paper is intended as a guide to this convening process for both the external professionals as well as the "internal" professionals they will engage and who will utilize the outcomes of the process.

The paper is in two parts. First, a review of the convening process will be presented. This will encompass both the generic aspects of the process as well as some specifics relevant to the case at hand. Second, a brief overview of the substantive issues involved and the objectives of this specific endeavor will be reviewed. The paper is intended as a briefing paper. Unanswered issues and questions should be directed with dispatch to the authors. The refinement of these objectives and goals will be an agenda item during the initial meeting of the conveners.

THE CONVENING PROCESS-BACKGROUND

The concept of the convening process grew out of the work of Norman Gold and his colleagues in the evaluation research group of the former National Institute of Education during the early 1980's. Its purpose was to bring together some of the rigor and formality of the technology of educational program evaluation with the real world of educational policy making. In a sense, it can be seen as a merging of insights on evaluation with insights on knowledge utilization. Put simply, the required niceties and rigors of evaluation research rarely meet the real world demands on policy makers for timely decisions informed by other sets of data beyond the "scientific." The convening process adapts to that reality by bringing greater structure and discipline to the process of collegial consultation which is the most common form of "data gathering" in which many policy makers engage. In the words of Mr. Gold:

"For inquires limited both in terms of substance and time, a simple and focused process in needed. Since the problems a school system may encounter are not likely to be unique, experience exists for both facing and dealing with these problems. To gain this knowledge, one needs to identify colleagues who have been in similar circumstances and determine what worked and didn't work for them. Administrators and other decision makers frequently call on trusted colleagues or others they may have heard about for advice. The current interest in networks, professional meetings and other forms of collegial interchange reflects this purpose. This system appears to be quite influential in informing local decision makers. No set of individuals has greater credibility in a school system than other school people.

"The Convening Activity, capitalizes on this natural system of support, attempting to make it more useful and reliable. The process employed to accomplish this is designed to formalize the use of collegial consultation as an assessment and problem solving tool. The formalization is designed to facilitate: (1) access to relevant colleagues, (2) opportunities for problem identification and (3) the process for problem solution.

"This process is intended to bring together people who are representative of the range of experience and insight available in schools concerning the problem being addressed. The commitment to a set of achievable objectives for the consultation, augmented by the analysis of existing data and the on-site review, appreciably facilitates problem identification. Finally, employing a group process for problem solution is intended to yield recommendations more in line with a district's needs than a series of individual consultations might bring."

AN OPERATIONAL PLAN FOR THE BEGINNING TEACHER TESTING AND ASSESSMENT CONVENING PROCESS

The full scope of activities surrounding a convening event comprise five major elements or tasks. These include:

1. Determining a need for the process and developing its operational plan;

2. Collecting descriptive and analytic information including a) problem specification, b) identification of locations where similar problems have been engaged, c) setting of objectives for the process and d) orienting and informing participants;

3. Identifying and seeking agreement from the "external" colleagues who will participate;

4. Conducting convening activities including a) prior off-site review of materials, specification of questions and development of preliminary recommendations and b) on-site activities including the orientation session, the on-site review and interaction and on-site development of findings and recommendations; and

5. The presentation of results including the development of the group's report and its presentation to the district.

The balance of this section will review each of these tasks as they pertain to this specific convening event. In this section substantive information will be kept to the minimum necessary for understanding the process. The following section will describe in greater depth the substantive issues to be engaged.

TASK 1 Determining a Need for the Process and Developing its Operational Plan.

The District of Columbia Public Schools is unique in that its regular fiscal budget, and therefore its programs, are subject to the approval of the U.S. Congress in addition to the D.C. City Council and Mayor. In 1983 Congress charged the district with developing a "balanced and comprehensive system which will embrace the concepts of merit pay and exemplary teacher recognition...(in order) to recognize outstanding teachers through both monetary and professional incentives." Shortly thereafter the superintendent appointed a task force to develop a plan that

1. Improves the quality of the teaching environment;

2. Provides recognition and incentives for professional self development of teachers; and

3. Enhances the District's ability to attract and retain highly qualified teachers.

The task force commissioned several pieces of research which culminated in a May 1984 report entitled a "Study of Teacher Incentives in the District of Columbia Public Schools."

Among the recommendations of the report was a proposal to establish an Intern-Mentor program. The program was to be a new way of "selecting and inducting new teachers in the D.C. Public Schools...to provide more rigorous screening and more intensive support."

In December 1984 the Board of Education endorsed the Intern-Mentor Program concept and mandated that all new teachers entering the school system in 1985-86 be required to serve a year-long internship under the guidance of an experienced mentor teacher. An important element of the Intern-Mentor Program is a requirement that interns take a subject matter test in their teaching field. The Board of Education intends that such tests will be used in the process of hiring all new teachers.

Since the Intern-Mentor Program's inception that Program has had the responsibility for developing the District's teacher testing capacity. To date twelve subject area tests have been developed along with a writing assessment. These tests have been developed under contract with National Evaluation Systems, Inc.

DCPS now wishes to assess its current teacher testing capacity, compare it to the state of the art and the capacity of other jurisdictions and develop a teacher testing policy which is appropriate to its context and can serve as a viable guide to its use of teacher tests.

The office with responsibility for developing the Intern-Mentor Program and subject area tests is the Office of Incentive Programs. The office is directed by Joan Brown who served as Assistant to the superintendent for the study of merit pay for teachers. The Superintendent has asked the Division of Quality Assurance and Management Planning, (DQA) directed by Mr. David Huie to assist in advising the Board on the development of a techer testing policy. Mr. Huie has turned to Mr. Norman Gold, Director of the Office of Research and Evaluation and the original developer of the convening process, to take responsibility for applying the convening model to this matter. Mr. Gold will be assisted in this task by three consulting members of his office, Mr. Dennis Holmes, Ms. Barbara J. Williams and Mr. Michael Kane. These individual have extensive experience with test development and teacher policy issues. Mr. Holmes and Mr. Kane have also conducted a convening event for DCPS previously. The balance of this section describes the operational plan for the convening event.

Task 2 Collecting Descriptive and Analytic Information

Subtask A Problem Identification

The statement above has described the general problem area this process is to engage. The substantive section of this briefing paper will specify this problem in greater detail along with objectives for the process. Here we will briefly describe the process for specifying the problem statement.

Ms. Brown and the various DQA staff involved have had several general discussions of the professional development

issues faced by the District. These discussions have been amplified by a review of the 1984 task force report and the reports on subject area test development. This briefing paper, which includes a problem statement, was written by Mr. Kane and Ms. Williams as a synthesis of discussion held to date. This draft has been reviewed by DCPS staff and the problem statement has been refined to incorporate their additional input. The colleague consultants will receive this draft.

Subtask B Identification of Locations Where Similar Problems Have Been Engaged.

Teacher Testing has been a topic of increasing attention since the early 1980s. Attention to it was reinforced by the 1983 Nation at Risk report. At present, approximately 44 states have some type of testing required for full teacher certification. With this relative wealth of experience to draw upon, our goal in selecting consultants to advise DCPS was focused upon finding those whose testing experience most closely matched the particular presenting issues in DCPS.

Accordingly, a search has been conducted for consultants who are familiar with the specific types of customized tests being used in D.C., who have high levels of technical expertise in test construction, who are expert in the legal implications of teacher testing, and who have been involved with the issues of the testing performance of minority persons. We also sought participants who were from states that drew upon the same or similiar pools of prospective teachers.

As regards the testing of beginning teachers, DCPS acts more as an SEA than as a school district. Therefore states were seen as a primary source of expertise. However, school districts were consulted to determine if their experience with the issue offered a unique perspective that should be included. This was determined not to be a critical perspective in this situation. Accordingly, the convening consultants are drawn primarily from SEAs and from those higher education institutions that advise them.

Subtask C Setting Objectives

The setting of objectives followed the same process described for Subtask A, problem identification. The convening process is more formal than day-to-day collegial consultation. This formalization provides structure which improves the usefulness and reliability of the consultation. For that reason (i.e., to maximize the probability that the process will be productive) the objectives are being specified early in the process and will again be a subject of review in the initial meeting of the consultants on-site. Every attempt has been made to keep the objectives realistic and to deal with both problem

clarification and problem solving issues.

Subtask D Orientation and Information Phase

It is essential that in the Orientation and Information Phase the consultants become as informed about the District and its issues as possible. Knowledge about the current state of the District in dealing with the problem to be addressed is essential to the analysis of the consultants. The orientation and information phase has two components, one that precedes the actual convening of collegial consultants, and one that commences at the time of convening.

The Orientation and Information Phase commences after the objectives and the general strategy for achieving those objective have been agreed upon. This briefing paper and accompanying briefing book are intended to supply the convening group with a full-range of input on:

1. The problem as the District perceives it;

2. The range of differences in perception of the problem;

3. Alternatives already attempted to deal with the issue;

4. Description of the current system and state of implementation; and

5. Current thinking on what to do about the problem.

These issues should guide the group's review of documents before their arrival on-site.

The group chairmen have worked with DCPS personnel to compile information for the selected collegial consultants. This information is critical for the adequate preparation of the consultants. The more the consultants know about the history of the problem and of efforts to deal with it by DCPS, the more prepared they will be when providing their analysis and consultation. The consultants are receiving copies of the following documents:

1. NES Final Report

2. Writing Test Field Test Report

3. Examinee Manual and Lists of Learning Objectives Covered by Tests.

4. Parents United Report (an external citizens group) The Recruitment and Retention of Excellent Teachers

5. Teacher Incentives - Report of the Ad Hoc Committe on

the Study of Incentives

6. Career Ladders for Teachers, (an example of a "convening process " outcome).

7. The Report (Superintendent's Report to the B.O.E.)

8. Brochures on Intern Mentor Program, Teacher Tests, and Teacher Certification.

This information is not synthesized. The consultants are asked to spend 1 day reviewing it prior to arrival in Washington.

In addition to working with district officials to develop this reading list, the group chairmen will work with district staff to identify those individuals who should be interviewed on-site by the consultants. We anticipate those interviews will involve personnel from the superintendent to individual teachers.

Task 3 Identifying and Securing Agreement from the "External" Colleagues Who Will Participate

As discussed under Task 2 Subtask B above, the goal in identifying and selecting colleagues for the consultation was to develop a pool of candidates which included those with experience in states that had developed customized teacher tests as well as others who are involved in a broader array of test related activities, settings and issues. Using a snowball sampling technique approximately 20 colleagues who, taken together, met this criterion were contacted. Agreements to participate have been obtained from the following individuals.

Mr. David Boyd, Attorney at Law and Partner, Balch and Bingham, Montgomery, Alabama. Mr. Boyd's law practice focusses on the defense of claims arising under the Constitution, Civil Rights Acts and Voting Rights Acts. He has defended several educational institutions in cases involving testing and most recently took the Alabama teacher testing case to trial.

Mr. Ronald Braithwaite, Professor, Medical College of Hampton Roads Mr. Braithwaite has had extensive involvement in the use of tests and other assessment devices for student counseling in higher education settings as well as considerable professional involvement with analysing the impact of such devices on minority performance and participation in educational and employment opportunities. He was involved in conceptualizing the Virginia Beginning Teacher Assessment Program and in assessing the use of the NTE program.

Mrs. Vana Dabney, Supervisor of Educational Assessment, South Carolina Department of Education. Ms. Dabney has been extensively involved in both the development of customized subject area tests as well as with the validation of NTE exams

for use in South Carolina. She has reviewed customized tests developed by NES for use in South Carolina.

Mr. Thomas Fisher, Administrator, Assessment, Testing and Evaluation, Florida Department of Education. Mr. Fisher has been with the Florida DOE for over ten years and has led its extensive efforts in both teacher and student testing, including its successful implementation of the testing program which resulted in the seminal Debora P. Vs. Turlington legal case. In recent years he has been responsible for the review and validation of NTE exams as well as the customized development of subject area tests for both beginning and experienced teachers.

Mrs. Sylvia Johnson, Professor of Psycho Educational Studies, Howard University. As a Professor of tests and measurement at an historially Black institution located in the District of Columbia, Mrs. Johnson is uniquely qualified to assist DCPS assess its current teacher testing status and the likely impact of any changes on teacher recruitment.

Mr. William Mehrens, Distinguished Professor, Michigan State University. Mr. Mehrens has been involved in the development of numerous testing programs for both students and adults. He has advised on the development of several types of teacher tests including subject area tests and has served as an expert consultant and witness in litigation involving such tests. He is a former president of the National Council on Measurement in Education.

Mr. Lester Solomon, Director Teacher Assessment, Georgia Department of Education. Mr. Solomon has been responsible for the development of the extensive teacher testing program which has been established in Georgia. This program, which requires satisfactory performance on subject area tests for both initial as well as recertification purposes, uses tests which were developed by NES and which form the basis for the tests developed in the DCPS.

Task 4 Convening Activities

Subtask A Preliminary Off-Site Activities

The actual time available to collegial consultants is limited. We have estimated five days per consultant for all activities. Of the five days one will be spent in preparation for the site visit. The separate activities to be conducted during this phase can be broadly described as follows.

1. Review of Orientation and Synthesis Reports

   The materials described above as well as this briefing paper are being provided to the consultants in advance of their visit. This information consists of

significant and/or relevant documents to enable the consultants to gain as much insight in as short a period of time as possible.

2. Specification of Questions

As a result of their review of written materials, consultants will be able to specify questions they would like to address during their on-site interviews. These questions will be used to broaden the understanding of the consultants in their effort to help identify problems and offer suggestions for their solution. Specification of at least some questions in advance will help DCPS officials arrange interviews with the most knowledgeable school people. It is currently planned to interview groups of individuals representing teachers, the District's management officials, the community, the local teachers union and the Board of Education. Therefore, the consultants should develop questions for each of these groups prior to the time to be spent on-site.

3. Preliminary Attempt at Analysis and Recommendations

Prior to coming on-site it will be useful for consultants to go through an exercise for their own awareness. This consists of writing down preliminary recommendations based upon what was known at that point. This exercise will have two purposes. The first is to stimulate preliminary thinking about and organizing of information on the DCPS case. The second is to make explicit one's own position and set of biases which will be brought to the process.

Subtask B Convening the Group On-Site

The involved process for selecting collegial consultants, the setting of group problem solving objectives, and the provision of orientation and analytic information is all preparation for convening the group on-site. The group activities on-site will be conducted in three parts: (1) an orientation session; (2) an on-site review; (3) a structured group process leading to specific problem solving recommendations and a final discussion with relevant DCPS officials to present the preliminary findings and recommendations.

1. Orientation Session

The Orientation Session will occur during the evening prior to the first full day. Its purpose is to allow conveners to meet one another and key DCPS personnel. Plans will be gone over and the entire agenda along with logistics fully discussed. Preliminary questions will

be reviewed as preparation for the interviews to be conducted during day 1. Objectives for the event will be reviewed and, if necessary, refined.

2. The On-Site Review - Day 1

The purpose of the on-site review is to interview central actors responsible for solving the particular problems of the DCPS, as well as others in the system affected by the specific problems under consideration. The goal of this face-to-face interaction is to gain as much knowledge as possible directly from informants concerning the nature of the problem. This day is viewed as an extension of the orientation and information phase, therefore, the same general questions which guided the review of documents can guide this latter set of on-site information gathering activities (See Task 2 Subtask D).

The interviews will take place in one day. The consultants already will have considerable information concerning the problem being addressed from reports they have received and studied. They should have in mind exactly what they need from these interviews to complete their review. If the number of individuals or groups to be interviewed is too great for the time allotted, members of the convening group will split and conduct separate interviews. At the end of the day each interviewer will synthesize his or her notes and prepare general impressions. After this an evening meeting will be held to go over the interviews so that all members are familiar with the information gained from each session. In addition, this evening session will be used to develop tentative recommendations regarding actions to be taken by the DCPS.

3. The On-Site Convening Session - Day 2

The Convening Session itself will last one day and consist of two primary areas of discussion: (1) findings of the panel members, and (2) recommendations for the District. The product of the session will be an extensive outline of a report of findings and recommendations. The goal of the group session is to determine the group's collective findings and recommendations. The report is to reflect the sentiments of the group, not its individual members. Therefore, there will be a single group report, as opposed to individual members' statements. The group session will specify the outline of that report and writing assignments. During a part of the session the group will work individually or in smaller groups to develop an expanded outline of the report.

Though the report does not -- and probably should not -- present only one analysis or a single recommendation, the group will be required to reach consensus in support of the position they propose. This process of reaching agreement is intended to produce a more thoughtful, integrated and practical set of recommendations than could be obtained by any member individually. Positions presented by members can be challenged, modified or discarded in favor of positions the group decides are more useful for the District's needs. The process is designed to level off consultation from "try my way" to the adoption or adaptation of the group's experience.

When the convening consultants have an outline of their report and recommendations prepared, they will have an "exit interview" with the team of concerned District officials. At this interview the preliminary findings and recommendations will be presented. The group will discuss DCPS reactions and any necessary changes in the outline will be incorporated (or plans for their incorporation will be made) and the group will depart.

Task 5 The Presentation of Results

Once the group position has been formulated and outlined, the process of informing the school district will begin. The presentation of findings will be in two parts. As indicated above, the first will consist of an exit interview with appropriate school officials. This presentation will be, of course, preliminary. Its objective will be to present the group findings and recommendations as they are currently formulated. This discussion will also give receiving school officials the opportunity to ask questions and to react generally. Their feedback will be valuable for the development of the final report.

The second part, the draft final report, will be delivered to the school district within thirty working days from the exit interview. This time line will allow the chairmen to receive the individual writing assignments and to synthesize them into a brief, focused paper of findings and recommendations and to circulate it for review and comments to both the members and the District. If any member feels the report should make a statement not endorsed by the group, she or he may wish to write a minority position to be included in the final document.

OPERATIONAL PLAN
Time and Task Summary

TASKS	October	November	December
Task I - Conceptualization			
a) Conceptual Paper and Operating Plan			
b) Consultation/Review			
Task II - Developing Orientation and Analytic Information			
a) Problem Identification			
b) Criteria for Consultant Selection			
c) Preliminary Development of Objectives			
d) Identification of Orientation Materials			
Task III - Identification of Sites and Consultants			
a) Identification of Sites			
b) Acquisition of Consultants			
c) Finalizing Objectives for Convening Process			
Task IV - Convening Activities			
a) Material Review			
b) On-Site Activities			
Task V - Final Report			
a) Preparation			
b) Review			

MILESTONES

Share briefing paper and materials with consultants	November 6
Draft questions for interviewees sent by consultants	November 13
Convening Event	November 17, 18, 19
Report writing assignments due to chairmen from consultants	December 4
Draft final report due to consultants and District	December 17
Consultant Feedback on Draft Report Due Chairmen	January 2
Final report due District	January 8

BACKGROUND, STATEMENT OF THE PROBLEM AND OBJECTIVES FOR THE TEACHER TESTING CONVENING PROCESS

BACKGROUND

The District of Columbia Public Schools (DCPS) is a large urban school system serving a student population of about 85,000 in nearly 200 separate facilities. Students are predominantly black (96%) and lower income (85% are eligible for free or reduced price meals). The teacher population of 6,062 is primarily black, highly experienced and highly educated. The District's annual operating budget is $430,565,000. Per pupil expenditure during the 1986-87 school year was $4,297 and average teacher salary was $33,797. DCPS is unique in that its regular fiscal budget, and therefore its programs, are subject to the approval of the U.S. Congress and the President in addition to the D.C. City Council and Mayor.

The core of the school system's instructional program is its Competency Based Curriculum (CBC). The program was initiated in 1976 in response to concern about declines in student achievement. CBC is a skills mastery program geared to individual differences in learning style and rate of growth. The primary goal is to ensure that students acquire the skills and competencies necessary for successful functioning in adult roles.

During the school year 1980-1981, the D.C. Public Schools put into effect a comprehensive plan for monitoring student progress. The Student Progress Plan (SPP) is an integral part of the school system's Competency Based Curriculum. The primary purpose of SPP is to ensure students have acquired a satisfactory skills level before they are assigned to the next higher grade. The plan divides each traditional grade, 1 though 6, into two grade levels, A and B. Promotion and retention decisions are made at the end of each semester (January and June) of each school year. Currently, students in grades 1 through 6 are promoted if they have mastered at least 70% of skills, including all critical skills, required for their grade level in both reading and in mathematics. Students who have mastered the required skills in only one of these areas are promoted with transitional instructional status. Special instruction is then provided in the deficient subject area. Students whose skills mastery falls below the required level in both subject areas are retained.

The instructional programs in the junior high schools and high schools are characterized by an extension of the philosophy which guides the elementary school programs. A competency based curriculum has been implemented in grades 7-9. Next year this program will be extended to grades 10 and 11. This curriculum is complemented by a student tracking system which measures accomplishments in the basic skills and requires the provision of

remediation programs for those students deficient in these realms.

In addition to this joint curriculum and student tracking program the Comprehensive Secondary School Improvement Initiative requires the secondary schools to identify their needs and goals (in objectively measurable areas such as attendance, instructional remediation programs, school climate, etc.) and to implement specific efforts to achieve these goals. Schools are monitored according to their ability to deliver on their objectives.

Student performance in the elementary grades exceeds the national norm (CTBS Total Battery 60th percentile in grade 3 and 55th percentile in grade 6). However, grades 8 and 9 scores are slightly below national norms (46th and 47th percentile respectively). By grade 11, average scores have fallen to the 31st percentile level. An analysis of student cohorts indicates that respective national ranks fall as the cohort moves through the progressive years of schooling.

As a part of its ongoing program to improve the quality of instruction in District schools, and in response to the Congressional "merit pay" mandate described above, the DCPS conducted a study of teacher incentives during the 1983-1984 school year. The study was intended to "provide a comprehensive data base about teachers and teaching policies and conditions in the D.C. Public Schools." The convening consultants will have the summary of this highly competent study as background material therefore, it will not be reviewed extensively here. Essentially the study task force found:

- 75% of D.C. teachers will be either eligible to retire or will have actually retired by about 1993 thus creating a significant opportunity to affect teacher quality through recruitment, selection, and induction policies.
- D.C. teacher salaries seem to be competitive with other school districts but not with other occupations prospective teachers might also consider entering.
- DCPS appears to be lacking in well developed mechanisms to provide supervision and assistance to new teachers during the years the highest attrition rates from the profession are found.
- D.C. teachers do not have many opportunities to assume differentiated roles or responsibilities for additional pay.
- D.C. teachers tend to be supportive of existing (but minimal) awards programs and desirous of additional

opportunities for extra pay and responsibility as long as it's not merit pay.

Accordingly, the Task Force recommended:

- Creation of a mentor teacher program to assist in the induction of new teachers to full-time professional practice.

- Expansion of teacher incentive programs to provide opportunities for recognition and the dissemination of effective practices.

- Development of a school incentive award program to recognize outstanding school programs.

- Long range planning for a more comprehensive career ladder for teachers. This career ladder would link major salary increases to performance based career advancement and would restructure induction and promotion practices.

Since the report was issued a mentor teacher program and a department chair program have been initiated. These are described in brochures included in consultant review materials. A teacher awards program has been expanded through several diverse activities and a school incentives award program is under consideration.

No comprehensive career ladder, differentiated staffing or merit pay program has yet been attempted. While the DCPS Board accepted the task force report, including the career ladder concept in principal, the District adopted a conscious strategy of first working to develop what might eventually become relevant elements of a career ladder program and then addressing the possibility of a comprehensive program. There is significant interest at top levels of the District in improving the professional role of teachers and in enhancing the involvement of teachers in professional aspects of the functioning of schools. This interest, together with the long standing commitment to the issue of career ladders, raised the question whether initiating the development of a district-wide career ladder/differentiated staffing program would facilitate such a shift in the role responsibilities of some or even all teachers and if so, how to best implement such a program. A convening event was held this spring to help to the District respond to this question. The report of that event has been included in the consultants' materials.

Although there was no career ladder program initiated, the recommended Intern - Mentor program was pursued with great vigor. This program is described in a brochure included with the briefing materials which should be reviewed in its entirety. As

is stated in that brochure:

> "An integral part of the Intern-Mentor Program is the requirement that interns take a subject matter test in their teaching field. The first two classes of interns comprise the pilot groups for test development. In subsequent years, the tests will be used as part of the screening process for hiring new teachers."

Just how the newly developed tests should be used as "part of the screening process for hiring new teachers" is the central question being put to the convening consultants.

STATEMENT OF THE PROBLEM

For the past three years content knowledge tests have been under development by DCPS. Their development and use to date have been the responsibility of one district program. Until recently little was known by other elements of the DCPS management structure about the process by which these tests have been developed and for what specific purpose they have been developed.

However, the DCPS Board of Education has expressed its intent to require content knowledge testing for all prospective teachers. The question of whether this testing is to be for diagnostic, licensure, hiring or tenure purposes seems to be at a finer level of detail than the Board has yet considered. It does, however, expect advice from the administration on this issue shortly.

Accordingly, the Superintendent of the District of Columbia Public Schools has convened a special task force composed of representatives of the Division of Human Resource Management, Division of Quality Assurance and the Office of Incentives Programs. This group is operating with a charge to prepare, by December, 1987 a set of implementation and policy recommendations to accompany the installation of a teacher testing program in DCPS. The convening process will provide significant input to the three primary areas of emphasis of the task force's work. These areas include:

1. review of the current status of the content knowledge tests which have been developed for the District by NES, Inc. This review is to include a qualitative assessment of the test development process and of the products.

2. development of as much information as is feasible about the actual and likely performance of DCPS current and prospective teachers on the tests as developed as well as about the criterion-related validity of the tests.

3. consideration of the policy, management and personnel

issues that may influence the implementation of a systemwide assessment program for new teachers, should such a program be adopted by the Board of Education.

During the past month Division of Quality Assurance staff have been examining the new tests, the process supporting their development, and the performance of interns on the initial test administrations. A summary of that analysis is included as Appendix A of this paper.

The convened consultants will review the information the district has developed on the tests, interview further those responsible for test development, those responsible for the management of DCPS, those responsible for setting policy and a representative group of those who will be impacted by new policy initiatives in this realm, including both teachers and concerned citizens. Using the first hand information gathered by this exercise, in concert with their own considerable experience with this issue, the consultants will advise the district on the three dimensions cited above.

Given the extensive experience of the consultants it is not necessary for this paper to describe the specific issues imbeded in the three areas of task force activity identified. It may be useful however to note that they fall into two primary realms: the concepual and the operational.

By conceptual we are refering primarily to those technical issues of standards and procedure which guide the development, use and legal acceptance of tests and other employment related procedures. These include, especially, issues of validity and bias. The consultants are expected to advise on the quality of the current tests as it relates to these conceptual dimensions.

The consultants will also be expected to advise on operational issues as well. The District intends to implement a teacher testing program. It faces issues such as how should the current tests be used, for what types of decisions, what other tests or assessment processes might be necessary, where should the administrative responsibility for the recommended procedures reside, what support systems are needed for those required take the tests. etc. In addition, the district must consider the influence such a program might have on its collective bargaining agreements, its ability to recruit sufficient numbers of new teachers and on other legal issues testing teachers may engender.

OBJECTIVES

The objectives of this convening event derive directly from the discussion above. Simply put, they are to consider the information gathered and report to DCPS the collective judgement of the colleague group concerning:

1. What would be an appropriate policy position regarding teacher testing and assessment given the particular context, status and directions of DCPS.

2. What is the current level of development of the extant DCPS teacher tests relative to the state-of the-art in these matters, and

3. What is the best recommended use DCPS should make of the currently developed tests given the answers to the two previous questions and how should DCPS go about implementing that purpose.

There are a host of related questions that must be addressed to meet these objectives. Some are raised earlier in this paper, some may be raised by the conveners themselves and some will be raised by District staff to be interviewed. The conveners are asked to send any additional general questions to the chairmen as soon as possible. They are also expected to supply specific questions to be asked of DCPS interviewees at the same time.

APPENDIX D

AGENDA AND INTERVIEWEES

AGENDA

Convening Process to Assess the Development of And Make Recommendations for the Implementation of a Beginning Teacher Testing Program in District of Columbia Public Schools

November 17-19, 1987

November 17, 1987

7:00 p.m. Introduction and Orientation (Sumner School Room 302 17th & M Street, NW)

Participants: Consultants, Core DCPS Staff

Activities:
- Introduce participants
- Review and Clarify Objectives
- Review Preliminary Analyses and Recommendations
- Review Procedures for next two days

Goals:
- To introduce participants
- To reach consensus regarding group goals and procedures
- To make explicit the individual preliminary assessments and recommendations

Expected Outcomes:
- List of succinct objectives
- List of initial recommendations

Leaders: Michael Kane, Barbara Williams

8:30 p.m. Review of Interviewing Activities

Participants: Consultants, DCPS Core Staff

Activities: Review data requirements from interviewees

Goal: To reach consensus on panel's data requirements and procedures for obtaining them through face to face interviews

	Expected Outcomes:	Shared goals for interviews Specification of interview procedures
	Leaders:	Michael Kane, Barbara William

November 18, 1987

8:30 a.m.	On-site Review	(Superintendent's Conference Room) 12th Floor Presidential Building 415 12th Street, N.W.)
	Participants:	Consultants, DCPS Core Staff and invited respondents
	Activities:	Group interviews of stakeholder groups to identify issues and problems regarding test development and use, instructional staffing and receptivity to alternative solutions
	Goal:	To develop comprehensive data base from which recommendations may be developed
	Expected Outcomes:	Enhanced awareness of issues, problems and constraints concerning further development of DCPS' Teacher Testing Program
	Leaders:	Michael Kane, Barbara Williams
8:30 - 9:30 a.m.	Central Administration	
	Andrew Jenkins Mary Hendrick George Margolies P. Gary Freeman	Joan Brown Ken Nickoles David Huie
9:45 - 10:30 a.m.	Board of Education Members	
	Bob Boyd	

10:30 -
11:30 a.m. Test Development History and Procedures

Barbara Williams

Jeanne Clayton

11:30 -
12:30 Teachers Union

William Simons

Community Members

Hazel Brown
Delabin
Rice-Thurston

12:30 - 1:30
Lunch

1:30 -
2:30 p.m. Regional Administrators/
Principals

Shelia Handy
Barbara Jackson
John Sparrow
Lucille Christian

Teachers

Toni Hill
Karen Webster
Michael Rice
Emily Nalven
Jeffrey Choppin
Celcia Bell Dove

2:30 -
3:30 p.m. Review of Tests

Jeanne Clayton

3:30 -

5:00 p.m. Sharing of Data and Discussion

5:00 p.m. Private time and Dinner

7:30 p.m. Development of Tentative Recommendations
(Sumner School Rm G-3)

Participants: Consultants, Core DCPS Staff

Activities: Synthesize findings
Develop tentative recommendations

Goals: To initiate development of consensus on presenting issues and recommended solutions

	Expected Outcomes:	Identification of range of perspectives on presenting issues and recommended solutions
	Leaders:	Michael Kane, Barbara Williams

November 19, 1987

(Sumner School
Room 302

8:30 a.m. Convening Session

Participants:	Consultants, DCPS Staff
Activities:	Determine findings and develop recommendations
Goals:	To determine group's findings and recommendations and format of report
Expected Outcomes:	Specification of findings and recommendations keyed to objectives Specification of report's major topics
Leaders:	Michael Kane, Barbara Williams

11:45 a.m. Lunch

12:45 p.m. Outline Report and Specify Writing Assignments

Participants:	Consultants and core DCPS Staff
Activities:	Develop detailed outline of report and assignment of writing responsibilities
Goals:	To outline final report in detail
Expected Outcomes:	Report outline Writing Assignments
Leaders:	Michael Kane, Barbara Williams

3:30 p.m.	Exit Interview	(Superintendent's Conference Room)
	Participants:	Consultants, Core DCPS Staff, Senior DCPS Administrators
	Activities:	Present and discuss group's findings and recommendations
	Goals:	To communicate and refine findings and recommendations
	Expected Outcomes:	DCPS personnel aware of group's initial findings and recommendations Further refinement of findings, recommendations and report outline
	Leaders:	Norman Gold, Dennis Holmes
5:00 p.m.	Adjournment	

Convening Process

Career Ladder/Differentiated Staffing

AGENDA

WEDNESDAY, NOVEMBER 17, 1987

SUPERINTENDENT'S CONFERENCE ROOM
415 12TH STREET, N.W., 12TH Floor

SESSION	PARTICIPANTS
8:30 A.M. - 9:30 A.M.	Andrew Jenkins Deputy Superintendent
	Mary Hendrick Director, Personnel Certification and Accrediation
	George Margolies Legal Counsel, Legal, Regulatory & Legislative Branch
	P. Gary Freeman Director, Human Resource Management
	Joan Brown Director, Incentive Program for Teachers
	David Huie Director, Management Planning and Quality Assurance
	Ken Nickoles Director, Labor Relations Branch
9:45 - 10:30	Bob Boyd Board Member, Ward 6

10:30 - 11:30	Test Development Review Barbara Williams - Consultant, DCPS Jeanne Clayton - NES, Inc.	
11:30 - 12:30	Small Group Sessions With: Mr. William Simons President, Washington Teachers Union Ms. Delabian Rice-Thurston President, Parents United Ms. Hazel L. Brown President, D.C. Congress of PTAs	
12:30 - 1:30	Lunch	
1:30 - 2:30	Split Group With:	
	Administrators:	Teachers:
	Shelia Handy Region B Superintendent	Toni Hill Karen Webster
	Barbara Jackson Region C Superintendent	Michael Rice Emily Nalven
	John Sparrow Principal Harrison Elementary	Jeffrey Choppin Cecelia Dove-Ball
	Lucille Christian Principal Woodson Senior H.S.	
2:30 - 3:30	Review of Test Instruments Barbara Williams - Consultant, DCPS Jeanne Clayton - NES, Inc.	

APPENDIX E

SAMPLE REVALIDATION PROCEEDURES AND FORMS

DIRECTIONS FOR _*_ JOB RELEVANCE PANELISTS

As you know, one of the requirements for candidates for teacher certification in _____**_____ is that they pass a standardized test that covers knowledge and skills in their intended field of certification. For most areas of certification, specialty area tests of the _____*_____ are used. As part of the _____**_____ Board of Education's continuing review of state-conducted examination programs, you are asked to review an _*_ specialty area test. You are to judge the necessity of the knowledge and skills covered on the test for satisfactory performance as a beginning teacher in _____**_____. You are also to review the test items for potential bias.

More specifically, you will be asked to make two judgments regarding each item on two forms of the test. When you have completed the item-by-item review of the first form, you will be asked to make a total-test judgment for that test form. You will then make item-by-item judgments for the second form. After those item-by-item judgments, you will be asked to make a total-test judgment about the second form as well as a judgment about the equivalence of the two forms. Thus, your judgments will be made in the following order:

1. Item-by-item judgments for the first form
2. Total-test judgment for the first form
3. Item-by-item judgments for the second form
4. Total-test judgment for the second form
5. Form-equivalence judgment

* name of test
** name of state or district

The judgments you will make are described in greater detail below. Please make your judgments independently, without consulting other panelists.

Item-by-Item Judgments

You are to review each test item in the two test forms you will be given. For your reference, the correct answers for the items are provided on an answer key inside each test form. For each item, you are to answer two questions by marking your responses on the * Judgment Form provided.* Each of the questions is presented in a box below with an explanation of the question and guidance regarding how to respond. The questions are also printed at the top of each page of your * Judgment Form.

Job Relevance

> Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning **
> teacher in the specialty area tested?

As an experienced ** educator, you are in an excellent position to judge the necessity of the knowledge and skills tested in these * test items to the job require-

* There are some items that appear on both forms being reviewed. Because you will make judgments about these common items when you review the first form of the test, you need not make judgments about them when working on the second form. Thus, these items' numbers have been crossed out on the second half of your judgment form. It is important, however, that you reread these common items when reviewing the second form because you will be asked to make a total-test judgment about that form as well as about the comparability of the two forms.

ments faced by beginning _____**_____ teachers. In making your judgment about each item, think about the knowledge or skills that beginning teachers in the specialty area tested need to have in order to perform adequately on the job.

If you think that the knowledge or skill measured by a test item is necessary for satisfactory performance by a beginning _____**_____ teacher in the specialty area tested, then circle "Yes" for that item on the _*_ Judgment Form. If you think that the knowledge or skill being tested is not necessary for satisfactory performance by a beginning _**_ _____ teacher in the specialty area tested, then circle "No" for that item.

Bias

> Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socioeconomic status?

Note that this question deals with two aspects of bias by asking whether the item would "offend" or "unfairly penalize" any group of examinees. A test item that might offend certain examinees, for example, would be an item in which members of a group are portrayed in a stereotyped manner. To illustrate, if minority youths were depicted in an item as members of gangs while majority youths were not, then the item should be judged to be biased because it might offend minority examinees.

A test item that would unfairly penalize a particular group of examinees would be an item on which those examinees

perform less well than another group of examinees, even though both groups are at the same achievement level with respect to the knowledge or skill being tested. This difference in performance could be caused, for example, by dissimilar interests of the two groups. It could also be caused by differences in the two groups' mastery of a skill (or knowledge) irrelevant to that being tested. To illustrate, suppose a test item required examinees to draw a conclusion from a reading selection about a high school football game. It is possible that females would perform less well on such an item than males, not because they are less able to draw conclusions (the skill being tested), but because they may be less interested in and have less knowledge about football.

If you believe that there are elements in the test item that might offend or unfairly penalize any group of examinees, then circle "Yes" for this question. If you believe that the test item would not offend or unfairly penalize members of a particular group, then circle "No." If you answer "Yes," that is, if you believe an item might be biased, briefly explain (in the space provided or on the reverse side of the sheet) the nature of that bias.

Answer both questions for each item before proceeding to the next item. Please use a pencil. Be sure that your responses are marked in the space for the particular item being reviewed. Please make sure that all of your responses and comments are legible.

Total-Test Judgments

You will be asked to make two total-test judgments, once after completing your item-by-item judgments for the first form and again after reviewing the second form. For each form of the test, you are to provide an estimate of the

extent to which that test form's content is representative of the knowledge and skills needed by a beginning _____**_____ _____ teacher in the specialty area tested. To make this estimate, consider the full range of the subject-related knowledge and skills needed by a beginning teacher assigned to teach in the specialty area tested. After identifying, at least in general, the total set of subject-related knowledge and skills needed by such teachers, estimate the percentage of that domain of knowledge and skills that is covered on the test form you will have just reviewed. You will be asked to make that estimate by responding to the following question:

> What percentage of the subject-related knowledge and skills needed by a beginning _____**_____ teacher in the specialty area tested is covered on this test form?
>
> Estimated percentage = ____ %
> (to nearest five percent)

To answer this question, please supply an estimate, from 0% to 100%, that is rounded to the nearest five percent. To illustrate, if you believe that all, or almost all, of the subject-related knowledge and skills needed by beginning teachers in the specialty area tested are covered on the test form, then you might supply an estimate of 100% or 95%. If you believe, however, that only about half of the needed knowledge and skills are covered on the test form, then your response might be 50%. Clearly, your response will be only an approximation. By averaging all reviewers' estimates, however, we will be able to obtain a general idea of the

extent to which the test form's content is representative of the subject-related knowledge and skills needed by beginning teachers in the specialty area tested.

In making your judgment for the second form that you review, remember that some of the test items on the first form are also on the second form. Although you will not have judged the job relevance of these common items when reviewing the second form (having already done so during your review of the first form), you should take these items into consideration when you make your judgment about the second form's content.

Form-Equivalence Judgment

The final judgment that you will make concerns the equivalence of the content contained in the two test forms you will have reviewed. Although the difficulty of the two test forms can be determined on the basis of actual examinee responses, the comparability of the content covered by the two test forms must be established judgmentally. You will be asked to provide an estimate regarding the content-equivalence of the two forms by responding to the following question:

To what extent is the content of the two test forms comparable?

A. Almost Identical
B. Very Similar
C. Somewhat Similar
D. Very Different
E. Almost Completely Different

In making this judgment, focus on the categories of knowledge and skills that are represented on each form. For the two forms to be judged highly comparable, it is not necessary that individual items have identical content. Rather, base your judgment on the extent to which the same content categories are represented on the two forms. For example, suppose a mathematics test for high school students includes items from the category of solving word problems using the four basic arithmetic operations. Suppose one form of the test has word problems requiring students to multiply and add, and a second form includes word problems that require students to divide and subtract. Although the two sets of items are not identical, they both represent the content category of interest. Thus, on this content category, the two forms are highly comparable.

As when making the total-test judgment for the second form, be sure to take into consideration the common items (i.e., the items that appear on both test forms) when rendering your form-equivalence judgment.

SST JUDGMENT FORM

JOB RELEVANCE PANEL: SCHOOL PSYCHOLOGIST

Test Form: A

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
1	Yes No	Yes No	
2	Yes No	Yes No	
3	Yes No	Yes No	
4	Yes No	Yes No	
5	Yes No	Yes No	
6	Yes No	Yes No	
7	Yes No	Yes No	
8	Yes No	Yes No	
9	Yes No	Yes No	
10	Yes No	Yes No	
11	Yes No	Yes No	
12	Yes No	Yes No	
13	Yes No	Yes No	
14	Yes No	Yes No	
15	Yes No	Yes No	
16	Yes No	Yes No	
17	Yes No	Yes No	
18	Yes No	Yes No	
19	Yes No	Yes No	
20	Yes No	Yes No	
21	Yes No	Yes No	
22	Yes No	Yes No	

** name of state or district

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socioeconomic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
23	Yes No	Yes No	
24	Yes No	Yes No	
25	Yes No	Yes No	
26	Yes No	Yes No	
27	Yes No	Yes No	
28	Yes No	Yes No	
29	Yes No	Yes No	
30	Yes No	Yes No	
31	Yes No	Yes No	
32	Yes No	Yes No	
33	Yes No	Yes No	
34	Yes No	Yes No	
35	Yes No	Yes No	
36	Yes No	Yes No	
37	Yes No	Yes No	
38	Yes No	Yes No	
39	Yes No	Yes No	
40	Yes No	Yes No	
41	Yes No	Yes No	
42	Yes No	Yes No	
43	Yes No	Yes No	
44	Yes No	Yes No	
45	Yes No	Yes No	
46	Yes No	Yes No	
47	Yes No	Yes No	
48	Yes No	Yes No	
49	Yes No	Yes No	
50	Yes No	Yes No	
51	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
52	Yes No	Yes No	
53	Yes No	Yes No	
54	Yes No	Yes No	
55	Yes No	Yes No	
56	Yes No	Yes No	
57	Yes No	Yes No	
58	Yes No	Yes No	
59	Yes No	Yes No	
60	Yes No	Yes No	
61	Yes No	Yes No	
62	Yes No	Yes No	
63	Yes No	Yes No	
64	Yes No	Yes No	
65	Yes No	Yes No	
66	Yes No	Yes No	
67	Yes No	Yes No	
68	Yes No	Yes No	
69	Yes No	Yes No	
70	Yes No	Yes No	
71	Yes No	Yes No	
72	Yes No	Yes No	
73	Yes No	Yes No	
74	Yes No	Yes No	
75	Yes No	Yes No	
76	Yes No	Yes No	
77	Yes No	Yes No	
78	Yes No	Yes No	
79	Yes No	Yes No	
80	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
81	Yes No	Yes No	
82	Yes No	Yes No	
83	Yes No	Yes No	
84	Yes No	Yes No	
85	Yes No	Yes No	
86	Yes No	Yes No	
87	Yes No	Yes No	
88	Yes No	Yes No	
89	Yes No	Yes No	
90	Yes No	Yes No	
91	Yes No	Yes No	
92	Yes No	Yes No	
93	Yes No	Yes No	
94	Yes No	Yes No	
95	Yes No	Yes No	
96	Yes No	Yes No	
97	Yes No	Yes No	
98	Yes No	Yes No	
99	Yes No	Yes No	
100	Yes No	Yes No	
101	Yes No	Yes No	
102	Yes No	Yes No	
103	Yes No	Yes No	
104	Yes No	Yes No	
105	Yes No	Yes No	
106	Yes No	Yes No	
107	Yes No	Yes No	
108	Yes No	Yes No	
109	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
110	Yes No	Yes No	
111	Yes No	Yes No	
112	Yes No	Yes No	
113	Yes No	Yes No	
114	Yes No	Yes No	
115	Yes No	Yes No	
116	Yes No	Yes No	
117	Yes No	Yes No	
118	Yes No	Yes No	
119	Yes No	Yes No	
120	Yes No	Yes No	
121	Yes No	Yes No	
122	Yes No	Yes No	
123	Yes No	Yes No	
124	Yes No	Yes No	
125	Yes No	Yes No	
126	Yes No	Yes No	
127	Yes No	Yes No	
128	Yes No	Yes No	
129	Yes No	Yes No	
130	Yes No	Yes No	
131	Yes No	Yes No	
132	Yes No	Yes No	
133	Yes No	Yes No	
134	Yes No	Yes No	
135	Yes No	Yes No	
136	Yes No	Yes No	
137	Yes No	Yes No	
138	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?		Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?		Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
139	Yes	No	Yes	No	
140	Yes	No	Yes	No	
141	Yes	No	Yes	No	
142	Yes	No	Yes	No	
143	Yes	No	Yes	No	
144	Yes	No	Yes	No	
145	Yes	No	Yes	No	
146	Yes	No	Yes	No	
147	Yes	No	Yes	No	
148	Yes	No	Yes	No	
149	Yes	No	Yes	No	
150	Yes	No	Yes	No	

Total-Test Judgment: What percentage of the subject-related knowledge and skills needed by a beginning ** _____ school psychologist is covered on this test form?

Estimated percentage = ____ %
(to nearest five percent)

Test Form: B

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
~~1~~	~~Yes No~~	~~Yes No~~	
2	Yes No	Yes No	
3	Yes No	Yes No	
~~4~~	~~Yes No~~	~~Yes No~~	
~~5~~	~~Yes No~~	~~Yes No~~	
6	Yes No	Yes No	
7	Yes No	Yes No	
8	Yes No	Yes No	
9	Yes No	Yes No	
10	Yes No	Yes No	
11	Yes No	Yes No	
12	Yes No	Yes No	
13	Yes No	Yes No	
14	Yes No	Yes No	
15	Yes No	Yes No	
16	Yes No	Yes No	
17	Yes No	Yes No	
18	Yes No	Yes No	
19	Yes No	Yes No	
20	Yes No	Yes No	
21	Yes No	Yes No	
22	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socioeconomic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
23	Yes No	Yes No	
24	Yes No	Yes No	
25	Yes No	Yes No	
26	Yes No	Yes No	
27	Yes No	Yes No	
~~28~~	~~Yes No~~	~~Yes No~~	
29	Yes No	Yes No	
30	Yes No	Yes No	
31	Yes No	Yes No	
32	Yes No	Yes No	
33	Yes No	Yes No	
34	Yes No	Yes No	
~~35~~	Yes No	Yes No	
~~36~~	Yes No	Yes No	
~~37~~	Yes No	Yes No	
~~38~~	Yes No	Yes No	
~~39~~	Yes No	Yes No	
~~40~~	Yes No	Yes No	
~~41~~	Yes No	Yes No	
~~42~~	Yes No	Yes No	
~~43~~	Yes No	Yes No	
~~44~~	Yes No	Yes No	
45	Yes No	Yes No	
~~46~~	Yes No	Yes No	
~~47~~	Yes No	Yes No	
~~48~~	Yes No	Yes No	
~~49~~	Yes No	Yes No	
~~50~~	Yes No	Yes No	
~~51~~	Yes No	Yes No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** school psychologist?		Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socioeconomic status?		Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
~~52~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
53	Yes	No	Yes	No	
~~54~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
55	Yes	No	Yes	No	
56	Yes	No	Yes	No	
57	Yes	No	Yes	No	
58	Yes	No	Yes	No	
~~59~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
60	Yes	No	Yes	No	
~~61~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
62	Yes	No	Yes	No	
63	Yes	No	Yes	No	
~~64~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~65~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~66~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~67~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~68~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~69~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~70~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~71~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~72~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~73~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~74~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~75~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~76~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
77	Yes	No	Yes	No	
78	Yes	No	Yes	No	
79	Yes	No	Yes	No	
80	Yes	No	Yes	No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
81	Yes No	Yes No	
82	Yes No	Yes No	
83	Yes No	Yes No	
84	Yes No	Yes No	
85	Yes No	Yes No	
86	Yes No	Yes No	
87	Yes No	Yes No	
88	Yes No	Yes No	
89	Yes No	Yes No	
90	Yes No	Yes No	
91	Yes No	Yes No	
~~92~~	~~Yes No~~	~~Yes No~~	
~~93~~	~~Yes No~~	~~Yes No~~	
~~94~~	~~Yes No~~	~~Yes No~~	
~~95~~	~~Yes No~~	~~Yes No~~	
~~96~~	~~Yes No~~	~~Yes No~~	
~~97~~	~~Yes No~~	~~Yes No~~	
98	Yes No	Yes No	
99	Yes No	Yes No	
100	Yes No	Yes No	
101	Yes No	Yes No	
102	Yes No	Yes No	
103	Yes No	Yes No	
104	Yes No	Yes No	
105	Yes No	Yes No	
106	Yes No	Yes No	
~~107~~	~~Yes No~~	~~Yes No~~	
~~108~~	~~Yes No~~	~~Yes No~~	
~~109~~	~~Yes No~~	~~Yes No~~	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?		Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socioeconomic status?		Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
~~110~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~111~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~112~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~113~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
114	Yes	No	Yes	No	
~~115~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
~~116~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
117	Yes	No	Yes	No	
118	Yes	No	Yes	No	
119	Yes	No	Yes	No	
120	Yes	No	Yes	No	
121	Yes	No	Yes	No	
~~122~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
123	Yes	No	Yes	No	
124	Yes	No	Yes	No	
125	Yes	No	Yes	No	
126	Yes	No	Yes	No	
127	Yes	No	Yes	No	
128	Yes	No	Yes	No	
129	Yes	No	Yes	No	
130	Yes	No	Yes	No	
131	Yes	No	Yes	No	
~~132~~	~~Yes~~	~~No~~	~~Yes~~	~~No~~	
133	Yes	No	Yes	No	
134	Yes	No	Yes	No	
135	Yes	No	Yes	No	
136	Yes	No	Yes	No	
137	Yes	No	Yes	No	
138	Yes	No	Yes	No	

No.	Job Relevance Is the knowledge or skill needed to answer this item necessary for satisfactory performance by a beginning ** _____ school psychologist?	Bias Might this item offend or unfairly penalize any group of examinees on the basis of personal characteristics such as gender, ethnicity, religion, or socio-economic status?	Nature of Bias If you circled "Yes" to the bias question, please explain why you think the item might be biased. (Use the reverse side of this sheet if necessary, noting the number of the item to which your explanation refers.)
139	Yes No	Yes No	
140	Yes No	Yes No	
141	Yes No	Yes No	
142	Yes No	Yes No	
143	Yes No	Yes No	
144	Yes No	Yes No	
145	Yes No	Yes No	
146	Yes No	Yes No	
147	Yes No	Yes No	
148	Yes No	Yes No	
149	Yes No	Yes No	
150	Yes No	Yes No	

Total-Test Judgment: What percentage of the subject-related knowledge and skills needed by a beginning _____** _____ school psychologist is covered on this test form?

Estimated percentage = _____ %
(to nearest five percent)

Test-Form Equivalence: To what extent is the content of the two SST test forms comparable? (Check one.)

____ A. Almost Identical
____ B. Very Similar
____ C. Somewhat Similar
____ D. Very Different
____ E. Almost Completely Different

SST ITEM QUALITY FORM

Test _____

Form	Item	Comments
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____

SST ITEM QUALITY FORM

Test _____

Form	Item	Comments
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____

SST ITEM QUALITY FORM

Test _____

Form	Item	Comments
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____
____	____	_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

TASK I DIRECTION SHEET

PURPOSE:

To examine each Exit Examination Mathematics item and judge the performance of a hypothetical group of prospective high school graduates by using your experience and expertise. (This hypothetical group of prospective high school graduates is described below.)

MATERIALS NEEDED:

EXIT EXAMINATION MATHEMATICS TEST
TASK I DIRECTION SHEET
TASK I JUDGMENT FORM

DIRECTIONS:

Step 1 – Consider a group of prospective high school graduates who have achieved an adequate level of basic skills in mathematics for receipt of a South Carolina High School Diploma. Next consider a group who have not achieved an adequate level of basic skills in mathematics for receipt of a South Carolina High School Diploma. Now conceptualize that group of prospective graduates who are on the borderline between those who merit receipt of a South Carolina High School Diploma and those who do not. The performance of this hypothetical group of borderline prospective graduates represents the performance level that you must think about when making judgments for this task.

Step 2 – Carefully read Item 1 of the Exit Examination Mathematics Test and answer the questions as if you were taking the test.

Step 3 – Using your individual judgment, determine the percentage of the hypothetical borderline group that would be able to correctly answer Item 1.

Step 4 – Record your judgment, to the nearest five percent (5, 10, 15, etc.), for Item 1 in the space beside Item 1 on your Task I Judgment Form.

Step 5 – Repeat Steps 2–4 for each of the fifty (50) Exit Examination Mathematics items. **ALL ITEMS MUST BE COMPLETED.**

Step 6 – When you have completed your judgments, be sure to record the last four digits of your Social Security Number as your ID in the upper right hand corner of your TASK I JUDGMENT FORM. Also record your POSITION CODE:

A = Public School Administrator
T = Public School Teacher

Raise your hand and a staff person will collect your Task I Judgment Form.

ID_____
POSITION CODE_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

TASK I JUDGMENT FORM

ITEM	PERCENT	ITEM	PERCENT	ITEM	PERCENT
1	_____	21	_____	41	_____
2	_____	22	_____	42	_____
3	_____	23	_____	43	_____
4	_____	24	_____	44	_____
5	_____	25	_____	45	_____
6	_____	26	_____	46	_____
7	_____	27	_____	47	_____
8	_____	28	_____	48	_____
9	_____	29	_____	49	_____
10	_____	30	_____	50	_____
11	_____	31	_____		
12	_____	32	_____		
13	_____	33	_____		
14	_____	34	_____		
15	_____	35	_____		
16	_____	36	_____		
17	_____	37	_____		
18	_____	38	_____		
19	_____	39	_____		
20	_____	40	_____		

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

TASK II DIRECTION SHEET

PURPOSE:

To consider the Standard that has been calculated from your TASK I JUDGMENT FORM and, based on your experience and expertise, to revise the Standard, if desired, by considering your overall judgment of what the Standard should be.

MATERIALS NEEDED:

EXIT EXAMINATION MATHEMATICS TEST
TASK II DIRECTION SHEET
TASK II WORKSHEET

DIRECTIONS:

Step 1 – Locate your Standard in Box 1 on your TASK II WORKSHEET. It has been calculated for you from your TASK I JUDGMENT FORM.

Step 2 – Based on your knowledge of the test gained in TASK I and your knowledge of the context of the Exit Exam as outlined on Page 1, form an overall judgment of what the Standard should be. From this overall prospective, do you think your Standard from TASK I is too high, too low, or at the appropriate level?

Step 3 – If you are satisfied with your Standard from TASK I, record it in Box 2 on your TASK II WORKSHEET.

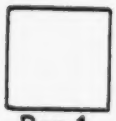
Step 4 – If you wish to reconsider your Standard, decide on a revised Standard. Record your revised Standard in Box 2 on your TASK II WORKSHEET. Keep your Task II WORKSHEET as you will need it in TASK III.

ID_____
POSITION CODE_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

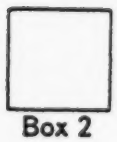
TASK II WORKSHEET

Your TASK I Judgments Resulted in a Standard of:



Box 1

TASK II Standard:



Box 2

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

TASK III DIRECTION SHEET

PURPOSE:

To consider a final revision in your Standard by examining the consequences with respect to the results from the Spring, 1986, administration of the Exit Examination.

MATERIALS NEEDED:

TASK II WORKSHEET
TASK III DIRECTION SHEET
TASK III WORKSHEET
DISTRIBUTION OF EXIT EXAMINATION SCORES TABLE

DIRECTIONS:

Step 1 – Find your Standard in Box 2 on your TASK II WORKSHEET. Record this number in Box 2 on your TASK III WORKSHEET.

Step 2 – Using the DISTRIBUTION OF EXIT EXAMINATION SCORES TABLE, locate your Standard in the column marked "Number of Items Correct."

Step 3 – Read across the row from your Standard to the column marked "Percentage of Examinees Who Would Meet the Standard." Record this percentage on your TASK III WORKSHEET in Box 3. This number represents the percentage of examinees expected to score at or above the Standard that you have proposed. The distribution tables are estimations based on examinees' performance on a 50–item Mathematics Test administered in April, 1986.

Step 4 – Read across the row from your Standard to the column marked "Percentage of Examinees Who Would Not Meet the Standard." Record this percentage on your TASK III WORKSHEET in Box 4. This number represents the percentage of examinees that is expected to score below the Standard that you have proposed.

Step 5 – The percentages in Box 3 and Box 4 of the TASK III WORKSHEET represent estimates of the percentages of prospective graduates who can be expected to either meet or not meet your Standard. Now consider these percentages in light of your knowledge of the context of the Exit Exam, your knowledge of the test from TASK I, and your overall judgment from TASK II.

a. Based on this information, if your Standard is acceptable, record your Final Standard in Box 5 on your TASK III WORKSHEET.

b. Based on this information, if your Standard is NOT acceptable, reconsider and revise your standard. Record your Final Standard in Box 5 on your TASK III WORKSHEET.

Step 6 – Be sure to record your ID and POSITION CODE in the upper right hand corner of your worksheet. Raise your hand and your materials will be collected.

DISTRIBUTION OF EXIT EXAMINATION SCORES
MATHEMATICS

Number of Items Correct	Percentage of Examinees Who Would Meet Standard	Fail to Meet Standard
4-10	100%	0%
11-13	99%	1%
14	98%	2%
15	97%	3%
16	96%	4%
17	96%	4%
18	94%	6%
19	93%	7%
20	91%	9%
21	90%	10%
22	88%	12%
23	86%	14%
24	84%	16%
25	81%	19%
26	79%	21%
27	76%	24%
28	73%	27%
29	71%	29%
30	68%	32%
31	64%	36%
32	61%	39%
33	57%	43%
34	53%	47%
35	49%	51%
36	47%	53%
37	42%	58%
38	39%	61%
39	35%	65%
40	32%	68%
41	28%	72%
42	24%	76%
43	21%	79%
44	17%	83%
45	14%	86%
46	10%	90%
47	7%	93%
48	4%	96%
49	2%	98%
50	0%	100%

ID_____
POSITION CODE_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September 17, 1986

TASK III WORKSHEET

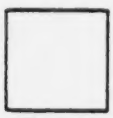
TASK II Standard:

Box 2

Percentage (%) of Examinees Who Would Meet Standard:

Box 3

Percentage (%) of Examinees Who Would Not Meet the Standard:



Box 4

Final Standard:

Box 5

EXIT EXAMINATION STANDARD SETTING COMMITTEE
MATHEMATICS
September, 1986

INDIVIDUAL RECOMMENDATION SHEET

Check one of the two boxes below:

☐ **Yes.** I accept the final vote as the committee recommended standard.

☐ **No.** I do not accept the final vote as the committee recommended standard.

If you checked "No," please indicate the score you would accept as the Exit Examination Mathematics Standard.

Comments:_____

_____

_____

_____

_____

_____

_____

_____
Signature (Optional)

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK I DIRECTION SHEET

PURPOSE:

To examine each Exit Examination Reading item and judge the performance of a hypothetical group of prospective high school graduates by using your experience and expertise. (This hypothetical group of prospective high school graduates is described below.)

MATERIALS NEEDED:

EXIT EXAMINATION READING TEST
TASK I DIRECTION SHEET
TASK I JUDGMENT FORM

DIRECTIONS:

Step 1 – Consider a group of prospective high school graduates who have achieved an adequate level of basic skills in reading for receipt of a South Carolina High School Diploma. Next consider a group who have not achieved an adequate level of basic skills in reading for receipt of a South Carolina High School Diploma. Now conceptualize that group of prospective graduates who are on the borderline between those who merit receipt of a South Carolina High School Diploma and those who do not. The performance of this hypothetical group of borderline prospective graduates represents the performance level that you must think about when making judgments for this task.

Step 2 – Carefully read Item 1 of the Exit Examination Reading Test and answer the questions as if you were taking the test.

Step 3 – Using your individual judgment, determine the percentage of the hypothetical borderline group that would be able to correctly answer Item 1.

Step 4 – Record your judgment, to the nearest five percent (5, 10, 15, etc.), for Item 1 in the space beside Item 1 on your Task I Judgment Form.

Step 5 – Repeat Steps 2–4 for each of the sixty (60) Exit Examination Reading items. **ALL ITEMS MUST BE COMPLETED.**

Step 6 – When you have completed your judgments, be sure to record the last four digits of your Social Security Number as your ID in the upper right hand corner of your TASK I JUDGMENT FORM. Also record your POSITION CODE:

A = Public School Administrator
T = Public School Teacher

Raise your hand and a staff person will collect your Task I Judgment Form.

ID_____
POSITION CODE_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK I JUDGMENT FORM

ITEM	PERCENT	ITEM	PERCENT	ITEM	PERCENT
1	_____	21	_____	41	_____
2	_____	22	_____	42	_____
3	_____	23	_____	43	_____
4	_____	24	_____	44	_____
5	_____	25	_____	45	_____
6	_____	26	_____	46	_____
7	_____	27	_____	47	_____
8	_____	28	_____	48	_____
9	_____	29	_____	49	_____
10	_____	30	_____	50	_____
11	_____	31	_____	51	_____
12	_____	32	_____	52	_____
13	_____	33	_____	53	_____
14	_____	34	_____	54	_____
15	_____	35	_____	55	_____
16	_____	36	_____	56	_____
17	_____	37	_____	57	_____
18	_____	38	_____	58	_____
19	_____	39	_____	59	_____
20	_____	40	_____	60	_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK II DIRECTION SHEET

PURPOSE:

To consider the Standard that has been calculated from your TASK I JUDGMENT FORM and, based on your experience and expertise, to revise the Standard, if desired, by considering your overall judgment of what the Standard should be.

MATERIALS NEEDED:

EXIT EXAMINATION READING TEST
TASK II DIRECTION SHEET
TASK II WORKSHEET

DIRECTIONS:

Step 1 – Locate your Standard in Box 1 on your TASK II WORKSHEET. It has been calculated for you from your TASK I JUDGMENT FORM.

Step 2 – Based on your knowledge of the test gained in TASK I and your knowledge of the context of the Exit Exam as outlined on Page 1, form an overall judgment of what the Standard should be. From this overall prospective, do you think your Standard from TASK I is too high, too low, or at the appropriate level?

Step 3 – If you are satisfied with your Standard from TASK I, record it in Box 2 on your TASK II WORKSHEET.

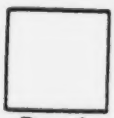
Step 4 – If you wish to reconsider your Standard, decide on a revised Standard. Record your revised Standard in Box 2 on your TASK II WORKSHEET. Keep your Task II WORKSHEET as you will need it in TASK III.

ID_____
POSITION CODE_____

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK II WORKSHEET

Your TASK I Judgments Resulted in a Standard of:



Box 1

TASK II Standard:

Box 2

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK III DIRECTION SHEET

PURPOSE:

To consider a final revision in your Standard by examining the consequences with respect to the results from the Spring, 1986, administration of the Exit Examination.

MATERIALS NEEDED:

TASK II WORKSHEET
TASK III DIRECTION SHEET
TASK III WORKSHEET
DISTRIBUTION OF EXIT EXAMINATION SCORES TABLE

DIRECTIONS:

Step 1 – Find your Standard in Box 2 on your TASK II WORKSHEET. Record this number in Box 2 on your TASK III WORKSHEET.

Step 2 – Using the DISTRIBUTION OF EXIT EXAMINATION SCORES TABLE, locate your Standard in the column marked "Number of Items Correct."

Step 3 – Read across the row from your Standard to the column marked "Percentage of Examinees Who Would Meet the Standard." Record this percentage on your TASK III WORKSHEET in Box 3. This number represents the percentage of examinees expected to score at or above the Standard that you have proposed. The distribution tables are estimations based on examinees' performance on a 60–item Reading Test administered in April, 1986.

Step 4 – Read across the row from your Standard to the column marked "Percentage of Examinees Who Would Not Meet the Standard." Record this percentage on your TASK III WORKSHEET in Box 4. This number represents the percentage of examinees that is expected to score below the Standard that you have proposed.

Step 5 – The percentages in Box 3 and Box 4 of the TASK III WORKSHEET represent estimates of the percentages of prospective graduates who can be expected to either meet or not meet your Standard. Now consider these percentages in light of your knowledge of the context of the Exit Exam, your knowledge of the test from TASK I, and your overall judgment from TASK II.

a. Based on this information, if your Standard is acceptable, record your Final Standard in Box 5 on your TASK III WORKSHEET.

b. Based on this information, if your Standard is NOT acceptable, reconsider and revise your standard. Record your Final Standard in Box 5 on your TASK III WORKSHEET.

Step 6 – Be sure to record your ID and POSITION CODE in the upper right hand corner of your worksheet. Raise your hand and your materials will be collected.

DISTRIBUTION OF EXIT EXAMINATION SCORES
READING

Number of Items Correct	Percentage of Examinees Who Would Meet Standard	Fail to Meet Standard
1-16	100%	0%
17-20	99%	1%
21	98%	2%
22	98%	2%
23	97%	3%
24	97%	3%
25	96%	4%
26	95%	5%
27	94%	6%
28	93%	7%
29	93%	7%
30	91%	9%
31	90%	10%
32	89%	11%
33	87%	13%
34	86%	14%
35	84%	16%
36	82%	18%
37	80%	20%
38	77%	23%
39	75%	25%
40	73%	27%
41	70%	30%
42	67%	33%
43	63%	37%
44	61%	39%
45	58%	42%
46	54%	46%
47	50%	50%
48	47%	53%
49	43%	57%
50	39%	61%
51	35%	65%
52	30%	70%
53	26%	74%
54	21%	79%
55	16%	84%
56	12%	88%
57	8%	92%
58	5%	95%
59	2%	98%
60	1%	99%

ID_____
POSITION CODE_____

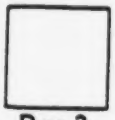
EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September 17, 1986

TASK III WORKSHEET

TASK II Standard:

Box 2

Percentage (%) of Examinees Who Would Meet Standard:

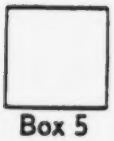


Box 3

Percentage (%) of Examinees Who Would Not Meet the Standard:

Box 4

Final Standard:



Box 5

EXIT EXAMINATION STANDARD SETTING COMMITTEE
READING
September, 1986

INDIVIDUAL RECOMMENDATION SHEET

Check one of the two boxes below:

☐ **Yes.** I accept the final vote as the committee recommended standard.

☐ **No.** I do not accept the final vote as the committee recommended standard.

If you checked "No," please indicate the score you would accept as the Exit Examination Reading Standard.

Comments: _____

_____

_____

_____

_____

_____

_____

_____
Signature (Optional)